A JOURNAL OF THE MILITARY
OPERATIONS RESEARCH SOCIETY
A JOURNAL OF THE MILITARY
OPERATIONS RESEARCH SOCIETY
A JOURNAL OF THE MILITARY
OPERATIONS RESEARCH SOCIETY
A JOURNAL OF THE MILITARY
OPERATIONS RESEARCH SOCIETY

Military Operations Research

Volume 2 Number 2
1996

A JOURNAL OF THE MILITARY
OPERATIONS RESEARCH SOCIETY
A JOURNAL OF THE MILITARY
OPERATIONS RESEARCH SOCIETY
A JOURNAL OF THE MILITARY
OPERATIONS RESEARCH SOCIETY
A JOURNAL OF THE MILITARY
OPERATIONS RESEARCH SOCIETY
A JOURNAL OF THE MILITARY
OPERATIONS RESEARCH SOCIETY
A JOURNAL OF THE MILITARY
OPERATIONS RESEARCH SOCIETY
A JOURNAL OF THE MILITARY
OPERATIONS RESEARCH SOCIETY
A JOURNAL OF THE MILITARY
OPERATIONS RESEARCH SOCIETY
A JOURNAL OF THE MILITARY
OPERATIONS RESEARCH SOCIETY
A JOURNAL OF THE MILITARY
OPERATIONS RESEARCH SOCIETY

MORS
MILITARY OPERATIONS RESEARCH SOCIETY

Military Operations Research

A publication of the Military Operations Research Society

The Military Operations Research Society is a professional society incorporated under the laws of Virginia. The Society conducts a classified symposium and several other meetings annually. It publishes proceedings, monographs, a quarterly bulletin, *PHALANX*, and a quarterly journal, *Military Operations Research*, for professional exchange and peer criticism among students, theoreticians, practitioners and users of military operations research. The Society does not make or advocate official policy nor does it attempt to influence the formation of policy. Matters discussed or statements made in the course of MORS symposia or printed in its publications represent the opinions of the authors and not the Society.

Military Operations Research, A Journal of the Military Operations Research Society (ISSN 1082-5983) is published quarterly by the Military Operations Research Society, 101 South Whiting Street, Suite 202, Alexandria, VA 22304-3418. The domestic subscription price is $40 for one year and $75 for two years; international rates are $80 for one year and $150 for two years. Application to Mail at Second-Class Postage Rates is Pending at Alexandria, VA.

POSTMASTER: Send address changes to *Military Operations Research*, A Journal of the Military Operations Research Society, 101 South Whiting Street, Suite 202, Alexandria, VA 22304.

Table of Contents

Volume 2, Number 2

1996

CALL FOR PAPERS

SPECIAL ISSUE ON

MILITARY OPERATIONS RESEARCH METHODS FOR FUTURE R&D CONCEPT EVALUATION

Limited research and development (R&D) budgets make it imperative that the United States analyze the potential operational benefits of future system concepts and select the most promising concepts for further R&D spending. Military operations research offers several techniques to help senior DoD decision-makers prioritize future R&D concepts. The purpose of this special issue is to describe the most effective techniques in use and to propose improvements to military R&D concept evaluation techniques.

Since many of the DoD R&D concept evaluation studies are classified, we are willing to publish the unclassified techniques or the techniques with notional data. However, in accordance with our editorial policy, we require certification from a senior decision-maker that the military R&D concept evaluation techniques were used.

Interested authors should submit **abstracts** by **January 15th 1997**. Papers should be submitted in accordance with our current editorial policy. All papers will be refereed.

We are also seeking volunteers to serve as guest editors, associate editors, and referees for this special issue.

Please contact me if you are interested in authoring a paper or serving as an editor/referee for this special issue.

DR. GREGORY S. PARNELL
Department of Mathematical Sciences
Virginia Commonwealth University
Oliver Hall, 1015 W. Main Street
P.O. Box 842014
Richmond, VA 23284-2014
Phone: 804-828-1301, ext. 133/Fax: 804-828-8785
Email: gparnell@vcu.edu and gsparnell@aol.com

Editorial Policy and Submission of Papers

Military Operations Research: What's Changed and What Hasn't?

Gregory S. Parnell, Editor, *Military Operations Research*

What hasn't changed?

We want to maintain and improve upon the high quality of articles published in *Military Operations Research.* The first Editor, Dr. **Peter Purdue**, and his Associate Editors have done a great job of carefully reviewing and selecting outstanding articles. The new editorial board and I will continue the high standards they have set.

What has changed?

The editorial policy has changed. We have developed procedures and instructions to authors that will expedite the review and publication process.

Our new editorial policy (see below) requests that authors identify the value of their analysis or research effort described in their paper. Authors must submit a statement of contribution and, for application articles, a letter from a decision-maker stating the benefits of the analysis or research.

The articles submitted to the journal cover many military operations research problem domains and methodologies. In order to assign the most appropriate reviewer, we have identified application areas and methodologies. We have also expanded the number of Associate Editors to insure we have expertise in all of these areas. In addition, we have developed procedures to insure timely review of submitted papers. To help expedite the publication process, we have developed instructions for *Military Operations Research* authors (see below).

EDITORIAL POLICY

The title of our journal is *Military Operations Research.* We are interested in publishing articles that describe *operations research* (OR) methodologies used in important *military* applications. We specifically invite papers that are significant military applications of OR methodologies. Of particular interest are papers that present case studies showing innovative OR applications, apply OR to major policy issues, introduce interesting new problem areas, highlight educational issues, and document the history of military OR. Papers should be readable with a level of mathematics appropriate for a master's program in OR.

All submissions must include a statement of the major contribution. For applications articles, authors are requested to submit a **letter** to the Editor—exerpts to be published with the paper—from a **senior decision-maker** (government or industry) stating the benefits received from the analysis described in the paper.

To facilitate the review process, authors are requested to categorize their articles by application area and OR methodology, as described by the following lists. Additional categories may be added. (We use the MORS working groups as our applications areas and our list of methodologies are those typically taught in most graduate programs.)

INSTRUCTIONS TO *MILITARY OPERATIONS RESEARCH* AUTHORS

The purpose of the "instructions to *Military Operations Research* authors" is to expedite the review and publication process. If you have any questions, please contact Mr. **Michael Cronin**, MORS Editorial Assistant (email: morsoffice@aol.com).

EDITORIAL POLICY AND SUBMISSION OF PAPERS

Composite Group	APPLICATION AREA
I. STRATEGIC	Strategic Operations
	Arms Control
	Revolution in Military Affairs
II. NAVAL WARFARE	Expeditionary Warfare/Power Projection Ashore
	Littoral Warfare/Regional Sea Control
III. AIRLAND CONTINGENCY OPERATIONS	Missile Defense
	NBC Defense
	Mobility
	Air Warfare
	Land Warfare
	Spec Ops/Ops other than War
	Air Defense
	EW & Countermeasures
	Joint Campaign Analysis
IV. SPACE/C3I	C3
	Mil Environmental Factors
	Oper Cont of Space Systems
	OR and Intelligence
V. RESEARCH & DEVELOPMENT	Measures of Effectiveness
	Test & Evaluation
	Unmanned Systems
	COEAs
	Weapon System Acquisition
VI. RESOURCES & READINESS	Soft Factors
	Social Science Methods
	Logistics
	Manpower & Personnel
	Resource Analysis & Forecast Readiness

OR METHODOLOGY
Deterministic Operations Research
Dynamic Programming
Inventory
Linear Programming
Multiobjective Optimization
Network Methods
Nonlinear Programming
Probabilistic Operations Research
Decision Analysis
Markov Processes
Reliability
Simulation
Stochastic Processes
Queuing Theory
Applied Statistics
Categorical Data Analysis
Forecasting/Time Series
Multivariate Analysis
Neural Networks
Nonparametric Statistics
Pattern Recognition
Response Surface Methodology
Others
Advanced Computing
Advanced Distributed Systems (DIS)
Cost Analysis
Wargaming

General

Authors should submit their manuscripts (3 copies) to:

Dr. Gregory S. Parnell, Editor, *Military Operations Research*
Military Operations Research Society
101 South Whiting Street, Suite 202
Alexandria, VA 22304

The manuscript should have camera ready illustrations and an electronic version of the manuscript prepared in WordPerfect or Microsoft Word. Per the editorial policy, please provide:

- authors statement of contribution (briefly describe the major contribution of the article)
- letter from senior decision-maker (application articles only)
- military OR application area(s)
- OR methodology (ies)

Length of Papers

Submissions will normally range from 5-25 pages (double spaced, 12 pitch, including illustrations). Exceptions will be made for applications articles submitted with a senior decision-maker letter signed by the Secretary of Defense.

Figures, Graphs and Charts

Please include camera-ready copies of all figures, graphs and charts. The figure should be of sufficient size for the reproduced letters and numbers to be legible. Each illustration must have a caption and a number which orders the placement of the illustration.

Mathematical and Symbolic Expressions

Authors should put mathematical and symbolic expressions in WordPerfect or Microsoft Word equations. Lengthy expressions should be avoided.

Approval of Release

All submissions must be unclassified and be accompanied by release statements where appropriate. By submitting a paper for review, an author certifies that the manuscript has been cleared for publication, is not copyrighted, has not been accepted for publication in any other publication, and is not under review elsewhere. All authors will be required to sign a copyright agreement with MORS.

Abbreviations and Acronyms

Abbreviations and acronyms (A&A) must be identified at their first appearance in the text. The abbreviation or acronym should follow in parentheses the first appearance of the full name. To help the general reader, authors should minimize their use of acronyms. A list of acronyms should be provided with the manuscript.

Footnotes

We do not use footnotes. Parenthetical material may be incorporated into a notes section at the end of the text, before the acknowledgment and references sections. Notes are designated by a superscript letter at the end of the sentence.

References

References should appear at the end of the paper and be unnumbered and listed in alphabetical order by the name of the first author.

POTENTIAL PAPERS OR SUGGESTIONS FOR THE JOURNAL

Military Operations Research is your journal. I need your help to identify the best articles for submission to the journal! If you have questions about a potential paper or suggestions for articles, please send me e-mail at gsparnell@aol.com.

I'm looking forward to seeing your article in *Military Operations Research*!

ANNOUNCING THE 65TH MILITARY OPERATIONS RESEARCH SOCIETY SYMPOSIUM

THEME:

ANALYSIS FOR COMPLEX, UNCERTAIN TIMES

Program Staff

Program Chair:
Dr. Harry Thie

Deputy Chair (Logistics):
Dr. Al Brandstein

Deputy Program Chair (Operations):
Kerry Kelley

Deputy Program Chair (Plans):
RADM Pierce Johnson, USNR

Site Coordinator:
COL Richard Linhart, USMC

Plenary/Special Sessions Coordinators
Ted Smith
Sue Iwanski
CAPT Lee Dick, USN
Anne Patenaude

Working Group/Composite Group Coordinators:
Jay Wilmeth
Bill Skeith
Bob Sheldon
LTC Mike McGinnis, USA
Mary Bonnet

Poster Session Coordinators:
Ken Martell
MAJ Willie MacFadden, USA
Gene Visco, FS

Tutorials Coordinators:
LTC Pat Vye
Dr. Dean Hartley
Phil Kubler

Prize Paper Session Coordinator:
Jim Duff

Education Session Coordinator:
Dr. Yupo Chan

VIP Coordinators:
MAJ Mark Gallagher, USAF

Junior/Senior Analyst Coordinators:
Robert Orlov
Howard Whitley
CDR Dennis Baer, USN

Spouse/Guest Coordinator:
Ginny Wiles

Marine Corps University, Quantico, VA
10-11-12 June 1997

Contact the MORS office for more information 703-751-7290 • E-mail: morsoffice@aol.com
http://www.msosa.mil.inter.net/mors

Executive Summaries

MODELING PRESENTATIONS AT NATIONAL CONVENTIONS

by Dean S. Hartley III

Organizing the presentations at professional society meetings is a chore that is common to all professions. The success of general purpose/national conventions is basically measured by the number of presentations that are obtained. Obtaining a maximum number of presentations with a minimum amount of work is a problem that lends itself to analysis using Operations Research. Dean Hartley analyzed the Military Applications Society contributions to national conventions and modeled them using Markov process concepts. The results are useful in minimizing the costs of recruiting session chairs for future meetings. The description of the analysis may also provide a useful educational tool.

STATISTICAL VALIDATION OF A COMMUNICATIONS NETWORK SIMULATION

by Ann E. M. Brodeen and Malcolm S. Taylor

Battlefield communications networks must deliver critical information when and where it is needed despite a rapidly changing and often hostile environment. Reliance upon computer simulations for system development and evaluation is often necessary since most communications systems are too complex to model analytically. Assurance that the simulation model faithfully emulates the process under study is essential in order to establish credibility and support the value of analyses and decisions that may follow. This paper describes a statistical procedure that provides an impartial assessment of agreement between simulated predictions and empirical observations for a communications network. The method is easy to understand, simple to implement, and flexible enough to hold the promise of more general and widespread application.

A DATA ANALYSIS OF SUCCESS IN OCS, THE USE OF ASVAB WAIVERS, AND RACE

by R.R. Read and L.R. Whitaker

The paper takes an in-depth look at the controversy posed by the facts that the use of ASVAB score waivers for admission to Officers Candidate School appears

i. unrelated to success in OCS when viewed by the individual races.
ii. related to success in OCS when the data are pooled into a single macro set.

The short answer is found in the fact that the use of waivers is quite variable from race to race. Further, increasing use of the waiver is associated with decreasing success rates in OCS. It is noted that the use of waivers diminished during the period of the study. The general result would be spurious if the OCS has some sort of racial bias internal to it. Another explanation is that the ASVAB or the administration of the decision rules has a bias that accepts candidates by race group, leading to uneven success rates in the school.

NON-MONOTONICITY, CHAOS AND COMBAT MODELS

by J.A. Dewar, J.J. Gillogly, M.L. Juncosa

While few combat modelers claim absolute predictivity for their models, many suggest their models are good at relative prediction—indicating when one system or configuration is better than another. Relatively predictivity requires a model to be monotonic in its outcomes—each additional increment of combat power for one combatant must lead to at least as good an outcome. This report shows that nonlinearities in a very simple deterministic combat models can produce non-monotonic results, where an additional increment of combat power leads to worse results. It further relates these non-monotonicities to the chaos that can bedevil nonlinear systems.

FINAL-COST ESTIMATES FOR RESEARCH & DEVELOPMENT PROGRAMS CONDITIONED ON REALIZED COSTS

by Mark Gallagher and David Lee

Managers and their analysts must estimate program costs and completion times. Most R&D programs, however, historically experienced significant schedule slips and incurred dramatic cost increases. Therefore, senior management wants risk assessment of on-going R&D programs. Gallagher and Lee propose a method that presents the probability of various final costs and completion times for an on-going R&D program. Since the approach relies on incurred costs, it avoids the problem of measuring against optimistic budget and schedule projections. Furthermore, while common methods only provide a point estimate, the proposed technique presents the likelihood the final program cost and completion time will be in any particular range. Program managers and their supervisors can use this information to assess the risk in continuing an R&D program.

Modeling Presentations at National Conventions

Dean S. Hartley III
Data Systems R&D Program

The contributions to national conventions, both in presenting work and in organizing sessions of presentations, are analyzed and modeled using Markov process concepts. The results are useful in minimizing the costs of recruiting session chairs for future meetings. The description of the process may also provide a useful educational tool.

The Military Applications Society (MAS) is one of the most active groups within either the Operations Research Society of America (ORSA) or The Institute of Management Science (TIMS). [ORSA and TIMS have merged into the Institute for Operations Research and Management Sciences (INFORMS).] MAS has sponsored up to three simultaneous full tracks at the ORSA/TIMS national meetings in good years and even in bad years manages at least one full track. Arranging the presentations requires hard work and perseverance and one would like to believe the results reflect that work. Despite the desire for credit when things go well, those who have been involved do admit that external factors play a role. The principal factor differentiating good times from bad appears to be the U.S. military budget. This is not surprising because changing budget levels affect the number of military applications that are funded (and that can be reported on) and changing budget levels often dramatically affect travel (for attending and reporting at conferences). As shown in Figure 1 and reported earlier (Hartley, *Operations Research*), the MAS presenters population is heavily concentrated around Washington, DC and in California. Thus the site of a conference may be an exogenous variable affecting presenters and outside the control of the MAS program organizer.

Beyond the three factors of budgets, site, and organizer's skill there remains a randomness to the response of potential presenters. Several modeling approaches come immediately to mind.

- One might regress some measure or set of measures of meeting success against these factors and then try to characterize the residuals.
- It has been noted (Hartley, *Phalanx*) that a large proportion of the presentations have been presented or arranged by a relatively small number of people (a general truism) and thus one might simulate the careers of "important" actors and characterize the residual activity in a "general" actor for the simulation.
- Either of these approaches might be worthwhile; however, I have chosen a third approach, one using concepts derived from Markov processes (Hillier and Lieberman).

THE CONCEPT

I keep track of contributions in a database. Each author of a paper gets a record (seven authors, seven records). Each session chair gets a record (and a session chair indicator). Each MAS meeting chair gets a record (and an indicator). I have another database with one record per person. (If I were doing this relationally, I would call these tables; but I'm not a purist.) In this latter database I keep a score for each person. I use a technique I learned back in the old

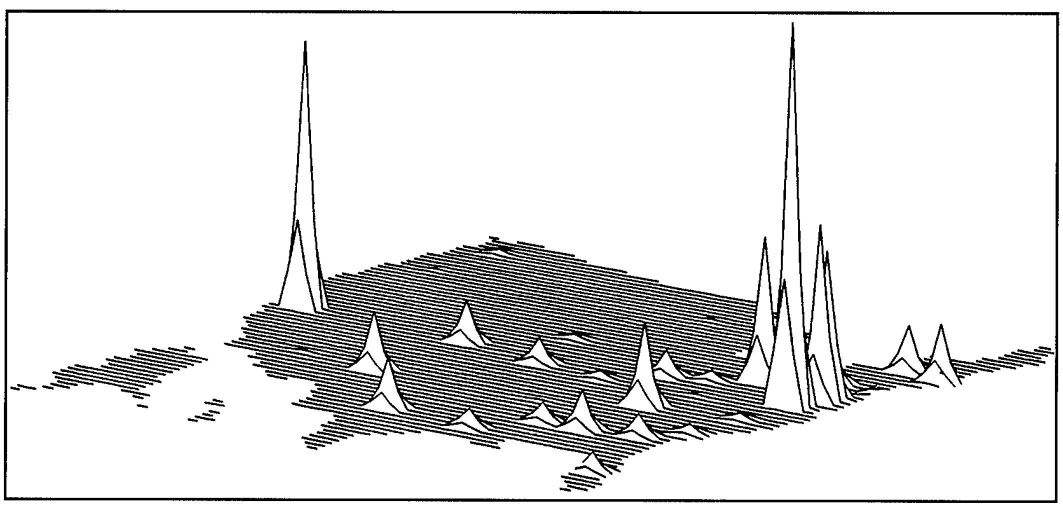


Figure 1. Concentrations of Contributors

Application Areas: Tutorial

OR Methodologies: Markov Processes

days (when computers were small, FORTRAN was new and you did packing by hand). Most people are authors only. I only need two decimal digits to count the number of papers in which they have participated. Many people (but not as many) chair sessions. If I multiply the number of sessions by 100 and add it to the number of papers, the two numbers don't overlap. I can read off either value with ease and still use only one word of memory. Finally, a very few people have been MAS meeting chairs. I multiply this value by 10,000 (because two digits for session chairs are sufficient) and add it to the score.

The reason for producing and maintaining this database was the expectation that it would make it easier to find people to create sessions for upcoming meetings. (Coincidentally, it makes the search for candidates for new officers less biased. The current officers can look at all those who are active contributors, not just those with whom they are personally familiar.) The intuitive concept was that someone who had already chaired a session was a reasonable candidate for chairing another session and someone who had presented three (or four) or more papers without chairing a session might be ready to chair a session. The scoring technique and the database make it trivial to print labels for all with scores greater than three or four. A later refinement allowed for retirement, death or loss of interest: only print labels for those whose last activity was after some date. (We also advertise generally to the membership in *Phalanx* and personally to acquaintances; but postage for solicitation letters costs money and the idea is to maximize the return while minimizing the cost.)

Those who are familiar with Markov processes have probably already guessed the punchline: each score is a state. A particular person will occupy a state, X, at a given meeting. At the next meeting the person may not be active (transitions back to the same state), may present a single paper (transitions to state $X + 1$), may chair a session (transitions to state $X + 100$), or may make multiple contributions (with n presentations and m sessions chaired transitions to state $X + m * 100 + n$). (A value of *10,000* may also be added to indicate chairing the MAS sessions; however, this will be ignored in the following discussion.) Questions of interest are: what are the transition probabilities and for each state, to what states are transitions likely? In particular, is the intuitive concept for recruiting session chairs valid and, if so, what is the proper cut-off score?

Students should be aware that the following discussion has been arranged for expository reasons in what I suppose to be a logical manner. The actual process proceeded using both iterative and parallel thought and computation processes. (The analyses were designed to answer or forward the study of more than one part of the investigation at a time.) I have found this to be the typical manner of real OR work. It is the responsibility of the researcher to dress up the results, putting them in an order accessible to others.

BUILDING THE MODEL

Before getting too deeply into the structure and parameters of the model, one aspect of the problem must be dealt with. Contributions to MAS sessions do not constitute a closed system. On the one hand, presentations are not restricted to a fixed population. Not only are young practitioners entering the arena of military applications all the time, but also people who normally work in other areas are free to present papers which are in some way associated with military applications of operations research. On the other hand, people are not required to continue in the field for ever: they may die; they may retire; they may continue to work in the field, but never present again; and they may simply leave the field. (For brevity, all reasons will be referred to as retirement.)

The Data

The data consist of more than 3400 records of contributions by almost 2000 different people to Military Applications sessions (sponsored and contributed) in 21 consecutive ORSA/TIMS national meetings. Several fields are contained in the database that are of no interest here, such as addresses, type application, and type methodology. The necessary items for this analysis are a personal identification field so that contributions over time can be tracked, the meeting identification, and the type contribution, whether presenting a paper or chairing a session.

First-Time Presenters and New People

Table 1 begins the analysis of the question of first-time presenters in MAS sessions. Simple sorting and counting is adequate to find the meeting in which each person in the database made his or her debut. The first column identifies the meeting by its date. The second column labels the meetings with a sequence number for alternative reference. The third column shows the result of the sorting and counting process.

The third column shows the number of people making their debut for each meeting per the database; however, this need not be the same as the first time that person ever contributed to a MAS session. The database starts with the May meeting in 1984, which was not the beginning of time. Because we are postulating differences in actions based on differences in state, this distinction may be important. Unfortunately, we have no way of obtaining the answer from the data—so we guess.

Column four shows the total number of different people contributing to each meeting (remember a person can make more than a single contribution at a meeting and for this purpose the contributions need to be collapsed). At the 84 May meeting there were 115 first-timers and 115 total contributors, which makes sense, as everyone there was presenting at the first meeting in the database. However, at the 94 April meeting, 61 people were first-timers of the total of 108 people.

Column five shows what percent the first-timers are of the total. The percentages for the first five meetings show a declining pattern and all are larger than the values for any of the rest of the meetings. We may assume that some of the first-timers are not "new people:" they have contributed at meetings previous to the start of the database. Because the sixth meeting has a percentage lower than any other meeting, I decided that this meeting was the first meeting in which the first-timers were identical to the new people. (The retirement analysis later lends some credence to this choice, as each first-timer who is not a new person must have a gap between his last presentation and his first in the database at least as large as the number of previous meetings in the database.)

The percentages in column five from meeting six through 21 vary, with an average of 67%. I used the 67% figure to estimate the new people at meetings one through five. The final column of the table shows these estimates, along with the assumed values for the rest of the meetings. Figure 2 shows the variation in the new input to MAS meetings graphically. The fact that this new input represents 67% of the contributors indicates the importance of new people to our meetings.

The graph appears to show a trend. Given the Reagan-Bush build-up of the military and the Bush-Clinton reductions, the existence of such a trend seems possible. However, it would be more convenient to the model if the variation represented the random variation of a sample

MEETING	SEQUENCE	FIRST-TIMERS	CONTRIBUTORS	PERCENT	NEW PEOPLE
84 May	1	115	115	100%	77
84 Nov	2	84	93	90%	62
85 Apr	3	103	123	84%	82
85 Nov	4	84	98	86%	65
86 Apr	5	76	100	76%	67
86 Oct	6	23	48	48%	23
87 May	7	155	203	76%	155
87 Oct	8	45	66	68%	45
88 Apr	9	176	237	74%	176
88 Oct	10	65	107	61%	65
89 May	11	80	138	58%	80
89 Oct	12	114	181	63%	114
90 May	13	121	189	64%	121
90 Oct	14	144	209	69%	144
91 May	15	88	130	68%	88
91 Nov	16	152	201	76%	152
92 Apr	17	75	129	58%	75
92 Nov	18	86	149	58%	86
93 May	19	61	97	63%	61
93 Oct	20	49	84	58%	49
94 Apr	21	61	108	56%	61

Table 1. First-Timers, Contributors and New People

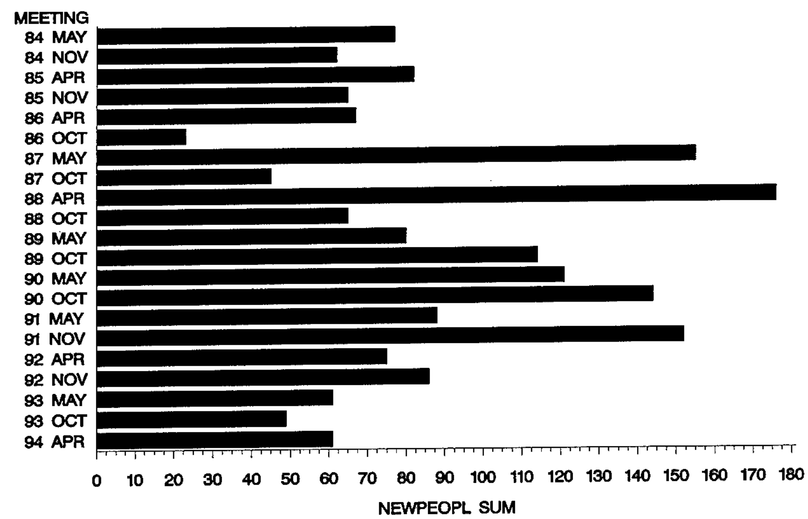


Figure 2. Contributions by New People in Each Meeting

drawn from a normal population. When one investigates tests for normality, one finds quite sophisticated tests and lots of room for error. It seems that determining (or rejecting) normality is very difficult.

Here I would like to assume non-normality and see if that assumption can be rejected; however, the tests I have work the other way - their null-hypothesis is one of normality. The analysis result is that I cannot reject the normality hypothesis at the 95% confidence level; but I can at the 90% level. Roughly stated, more than 5% of samples of the given size taken from a normal distribution will exhibit the observed abnormalities, but not 10%. If I were trying to reject normality, I would be faced with a potential error rate of more than 5% if I rejected normality. I would prefer a statistic that says X% of samples of the given size taken from the collection of all **non-normal** distributions will exhibit the observed normal characteristics. I would then have a statistic concerning my error rate for rejecting non-normality (and assuming normality). (Note that saying that 90-95% of samples of a given size from a normal distribution exhibit certain normal characteristics is not the same as saying that only 5-10% of samples of a given size from non-normal distributions will exhibit those normal characteristics. Thus, I am not 90+% sure of normality.) Because of my reversed worries, I cannot embrace the normal hypothesis. I must assume that accessions are driven by some non-random (or non-uniform random) process or processes.

Retirement from the Field of Military OR

Despite the problems of determining who is a new person, that determination is easier than determining that someone has permanently quit making contributions to MAS meetings. Not only do the data supply no clear markers for such an action, no determination can be positively made until after a person has died. Further, contributions can be made posthumously as a co-author. (Posthumous contributions will not support future requests to act as a session chair; however, such are the problems of the real world.)

One approaches this problem by approximation. If everyone presented in each consecutive meeting from their debut until their retirement, then one need only find a meeting in which a person did not contribute. The previous meeting would have been their last! While this supposition is contrary to fact, within its procedure lies the germ of a useable technique.

If a contributor presents at a meeting, skips the next, and presents at the next meeting, he or she has a gap of length one. Figure 3 shows a distribution of the lengths of the gaps of people

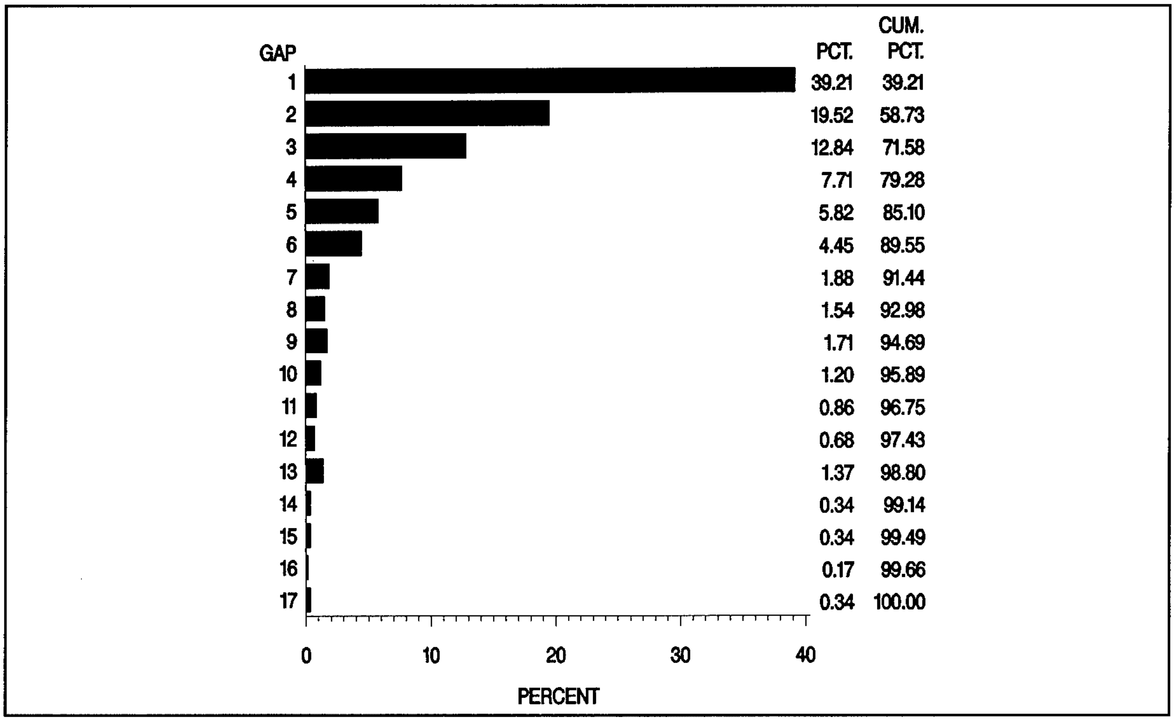


Figure 3. Distribution of Internal Gap Lengths

in the database. Only internal gaps are represented. These are gaps of known length because there are presentations at both ends. A first-timer has a gap that starts on or before the 84 May meeting; however, if the first-timer is a new person, this is not a gap. We have no way of distinguishing whether an initial gap is a real gap or not. Someone who retires has a gap ending in the database with the 94 April meeting; however, we have not figured out how to differentiate retirees from people who will contribute again. Internal gaps represent absences that are known to be temporary. Someone did not contribute for a while and then came back. We need an approximation that says if we see a gap greater than N extending to the end of the database, we may assume that that person retired after the last contribution.

From Figure 3, we see that if we use $N = 1$ and had cut off the database at various earlier dates we would have wrongly terminated careers 61% of the time. If we used $N = 17$, we would not have terminated any career in the database incorrectly; however, we have no guarantee that there will never be a gap greater than 17 that is internal. Further, we have only 21 meetings in the database, so that the longest terminal gap possible is 20. This value would imply very few retirement inferences.

Choosing $N = 6$ is more reasonable, as 90% of the internal gaps are of length six or less. Hence, if we see someone who has not contributed for seven meetings, we can be 90% sure that this person will not return. We can be even more confident that longer terminal gaps represent real retirement. This value will slightly overstate retirements for the 91 May and prior meetings. However, it risks great understatement of retirements for subsequent meetings, as no retirements at all will be inferred for these meetings. Accordingly, I have chosen $N = 5$, as a compromise between overstating and understating retirements.

What States Do New People Assume?

We do not know who all of the real new people are. We only know who all the first-timers are. Figure 4 shows the states (scores) and percents for each state that first-timers achieve in their first meeting. We will use the percentages

as probabilities. The chart shows that 85% of first-timers make one presentation at their first meeting. This is not surprising. The suspicion is that this percentage is artificially low. In the earlier meetings of the database, was it the first-timers who are not new people the ones starting off as session chairs and making multiple presentations? Because those people cannot be identified, this question can only be answered by inference. The technique is simple, divide the meetings into three segments: the first five meetings, the next eight meetings, and the final eight meetings. Then examine the distribution of scores for first-timers in each segment. The first timers in the latter two segments should be almost 100% new people.

The distributions for the latter two segments do appear more like each other than they are like the first segment. However, the differences do not appear significant in a practical way, given the level of resolution needed for this problem. The percentages for a score of one in the three segments are 84%, 86% and 85%, indicating close agreement. All three segments show strong contributions at the scores of 2, 100, and 101; however, the major difference is that the minor contributions at 200 and 201 are more pronounced in the first segment (just exceeding 1.0%, rather than just below 1.0%).

What Are the Transition Probabilities from each State?

Table 2 displays the transitions and percentages for nine states. The state is shown in a shaded rectangle at the upper left corner of its section of the table. The "0" state stands for the new people before their debuts. The bold numbers along the left side of each section are meant to be added to the bold numbers along the top of the section, as called out by each entry in the table. Thus the 85 in the first section has a one added to a zero, indicating that the zero state transitions to a score of one 85% of the time.

Reading the chart, one finds the transition probability from the "0" state to a score of "2" to be 3%, the probability of transition to a score of "100" to be 4%, and the probability of transition to a score of "101" to be 5%. The number "97" in the bottom row of the section shows that these probabilities sum to 97%. For clarity, only probabilities greater than 1% are shown. Thus 3% of

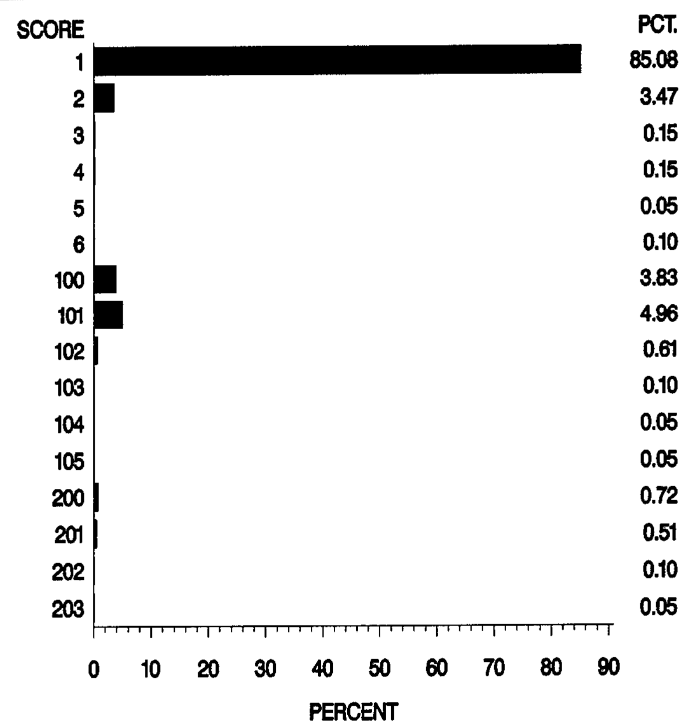


Figure 4. Distribution of States for First-Timers

0	**0**	**100**	**200**	**1**	**0**	**100**	**200**	**2**	**0**	**100**	**200**
-1				-1	28			-1	18		
0		4		0				0			
1	85	5		1	63			1			
2	3			2	7			2	65	1	
3				3				3	11	2	
4				4				4	1		
5				5				5			
6				6				6			
7				7				7			
8				8				8			
	97				98				99		
3	**0**	**100**	**200**	**4**	**0**	**100**	**200**	**100**	**100**	**200**	**300**
-1	20			-1	17			-1	24		
0				0				0	62	4	
1				1				1	5	2	
2				2				2			
3	48			3				3			
4	23	3		4	50	1		4			
5	3			5	17	2		5			
6				6	5	3		6			
7				7	2	1		7			
8				8	1			8			
	98				100				98		
101	**100**	**200**	**300**	**102**	**100**	**200**	**300**	**X**	**0**	**100**	**200**
-1	22			-1	20			-1	7		
0				0				0	64	3	
1	57	4		1				1	13	6	
2	7	6		2	61	2		2	2	2	
3	2	1		3	14	2		3			
4				4	2			4			
5				5				5			
6				6				6			
7				7				7			
8				8				8			
	99				100				96		

Table 2. Transitions and Probabilities

the transitions are omitted. These figures represent the same data shown in Figure 4.

Seven sections are devoted to the more typical states of a Markov process. The states "1" through "4" and "100" through "102" each have features lacking in the "0" state: they have transitions to themselves and transitions to the retirement state. The self transitions (e.g., state "1" to state "1") are shown to be 63%, 65%, 48%, 50%, 62%, 57%, and 61%, respectively. Self transitions lead to gaps, i.e., the contributor does not make a contribution in the succeeding meeting. His or her choice for the meeting after is independent of the choice (transition) made at this time. The convention used here is that the retirement state is the "-1" state. These transitions are 28%, 18%, 20%, 17%, 24%, 22%, and 20% respectively. These transitions are absorbing, taking that contributor out of the active population and leaving no succeeding choices. The percentages for these transitions are the percentage who never contribute again (where "never" is greater than five meetings and extends through the 94 Apr meeting). Each of these sections is derived from a distribution analogous to that of Figure 4, from which only transitions of 1% and greater are retained.

The underlying database of contributions is large, but not infinite. There are many states other than those with sections explicitly represented in Table 2; however, the number of transitions for these states are too small to produce reliable transition probabilities. Similarities grow as the states grow in numerical value. The similarities have a staggered pattern. Note the general decreases in retirement and self-transition probabilities among the single digit states and among the "100" states. However, the "100" states begin their sequence roughly at the "2" state position. The magnitude of the differences decreases as the numerical values grow. This

decrease in differences justifies the creation of a general state transition for all other states.

The general state is labeled the "X" state and is found in the ninth section of Table 2. This state represents all states not explicitly given in Table 2. It is created by adding all of the other state's relative transitions (relative transitions are described in next paragraph) and discovering their frequencies. The reason for doing this is that beyond some point there are too few examples of each state to create valid statistics. For example, if there were only three transitions from a given state, at most three successor states would be legitimated by statistics, whereas there would be no *a priori* reason to suspect transitions to other states to be illegitimate. Further justification for combining states is the observation that "experienced" presenters appear to have similar attitudes toward presenting, regardless of their exact level of "experience" (state value). Informal tests of the distributions were made to discover any reason to reject this hypothesis. No such evidence was found. The critical decision to be made concerns the identities of the states to be combined into the "X" state. This decision is based on size of each state and observed differences in transition patterns; however, it is somewhat arbitrary.

The "X" state section is arranged in a manner similar to the other sections. The "-1" state indicates retirement. However, the transition "to" states are found by adding the sum of the column and row positions to the value of "X." For example, if "X" represents the "411" state (four sessions chaired and 11 papers presented) the "6" in the body of the table represents a transition to one more session chaired (100) and one more paper presented (1), or a transition to the "512" state with probability 6%. The "X" state is similar to the "4" state and the "102" state in having several significant non-self-transitions; however, it has a much lower retirement transition probability. The retirement transition probability of 7%, as compared to the 17% to 28% for the other states, indicates that the people that reach the "X" state are committed to the field. The "X" state's 64% self-transition rate is in the high end of the range of self-transition values, indicating a slightly lower probability of repeating in successive meetings.

The model of the MAS contributor states and transitions is displayed graphically in Figure 5. Only those transitions given in Table 2 are displayed. (The picture would be unusably tangled, otherwise.) The semi-annual self-selection of new people as contributors (from Figure 2) is

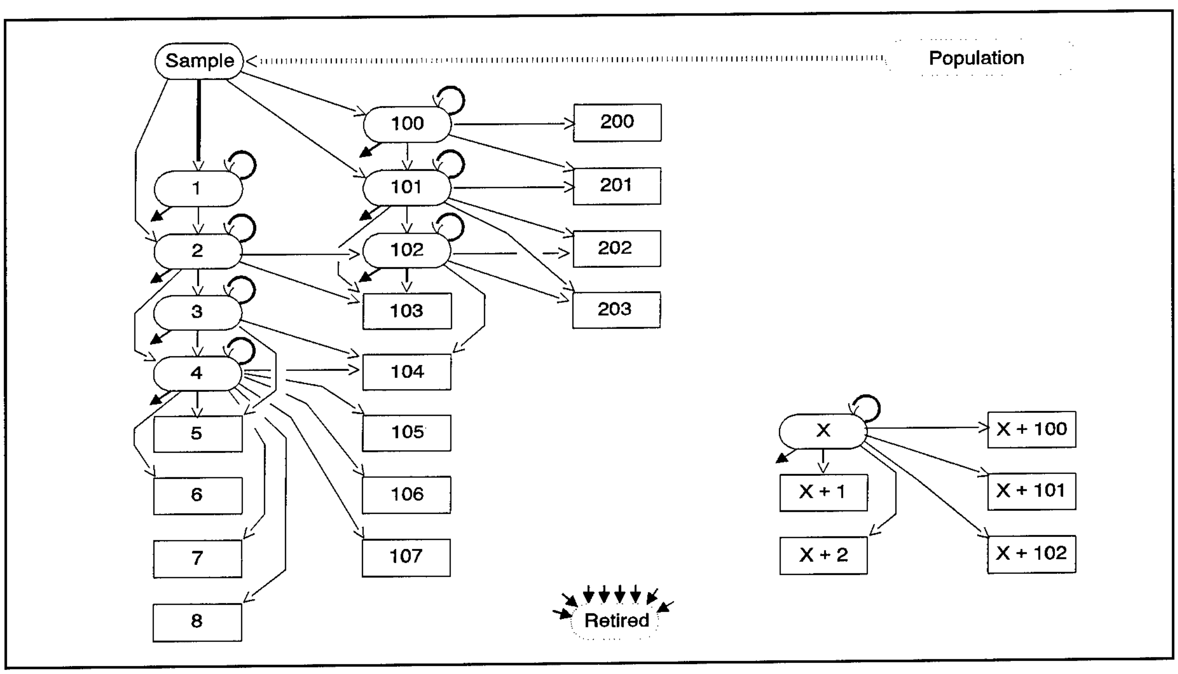


Figure 5. Model of MAS Meeting Contributions

represented by the dotted arrow from "Population" to "Sample" (the "0" state). The four major transition probabilities from the "0" state are shown with solid arrows. (The heavier arrows indicate higher probabilities.) Each of the states from which transition probabilities are explicitly given in Table 2 is shown as an oval with its score value inside. The looped arrow represents the self-transition. The short arrow that goes nowhere represents the retirement transition. (Each of these arrows logically connects to the "Retired" state; however, drawing the connections would needlessly complicate the diagram.) Each of the other states that is explicitly given as a "to" state is shown as a rectangle with its score value inside. Transitions from the states represented with rectangles and all other states are given by the disconnected "X" state transition diagram.

As described, the model is not explicitly finite. However, no one lives forever and the possible level of contribution at any given meeting by one person is finite. The implicit limit of a "9999" state imposed by the scoring process is certainly adequate. This limit could be made explicit by requiring that the only allowed transition from the "X" state for $X = 99nn$ or $X = nn99$ be the retirement transition.

Using the Model

Figure 5 is useful in seeing the flow of events. First-timers usually present one paper. Occasionally they present two or chair a session or chair a session and present a paper. (Note that the "100" state can only be filled by first-timers, just as the "1" state can only be filled by first-timers.) The lack of an arrow from the "1" state to any "100" state and the heavy arrows from the "2" state to the "3" state and from the "3" state to the "4" state indicate that the usual progression is more paper presentations, not chairing sessions. The "4" state also has a heavy arrow going to the next pure paper presentation state; however, it also has a significant spread of arrows to other states. This probably indicates diversification and maturity in the field.

The next observation illustrates a problem with a single view of the world: the diagram does not indicate where most of the "101" state people come from. The diagram indicates that less than one percent of those in the "1" state move to the "101" state (making their next appearance by chairing a session, but not presenting a paper) and that more than one percent of those in the "100" state do move to the "101" state. However, without information about the relative population of the "1" state and the "100" state, it is not clear that the flow from "100" to "101" is larger than the other. A diagram of "from" transitions could be constructed that would answer such questions. This situation does not present a problem here. Regardless of the number of session chairs deriving from the "1" state, the diagram does indicate that more than 100 letters would be required for each success and the total number of letters would be prohibitively expensive.

Figure 5 is also useful in determining who are the most likely recruits for new session chairs. All of those who have chaired sessions already (states "100" and above) show significant transition probabilities to chairing more sessions (adding 100 or more to the state). In addition, the people in the "4" state are clearly ripe for chairing a session. Also note that the states shown in rectangles participate in the "X" state transition diagram and should be regarded as the core contributors who will generate more papers and sessions in the future.

Further Uses of the Model[1]

A pure Markov model can be solved analytically. However, this model is somewhat impure because of its externally driven periodic input. Further, a solution that gives the probabilities for ending up in each absorbing state does not appear to translate into an interesting fact about the underlying reality.

There are, however, other interesting things one can do with a model defined using Markov state concepts. It is difficult to actually solve large Markov models. Even if the model is theoretically solvable analytically, it may be more practical to build a simulation to solve it. Such a simulation can answer additional questions that may be interesting in this context. For instance, what changes to the system will yield positive benefits? And what change directions maximize the result?

Obviously, these questions imply a metric. How does one measure goodness? And the questions imply knowledge of the current value of the measure. One such metric includes the number of currently active participants and their distribution over the states.

Table 3 shows the numbers (and percentages) in the significant states as of the April 1994 meeting. As before, the bold numbers in the left column represent numbers of presentations which are added to the bold numbers in the top row, representing numbers of sessions chaired. The total number of active participants is 639. Although this number represents what actually happened, it is not the only value that could have been realistically expected. A series of simulation runs could derive an estimate of the variance. (Note that the mean will be biased toward 639 because the transition probabilities were estimated using the historical outcomes.)

	0	100	X
0		11 / 1.7%	
1	391 / 61.2%	20 / 3.1%	
2	77 / 12.1%	12 / 1.9%	
3	14 / 2.2%		
4	7 / 1.1%		
X			107 / 16.7%

Table 3. State Occupancy by Currently Active People

Questions might be posed directly in terms of changes to the transition probabilities. Running sets of simulations with the changed probabilities would generate new results to compare against the baseline. Such a process would reveal some of the nature of the model; however, it does not reveal much about reality. Real options are likely to affect more than one transition probability. A major part of the problem lies in making reasonable guesses about the effects on transition probabilities of real options. For instance, suppose MAS were to award some nominal gift to some category of participant. Which transition probabilities might that practice raise, and by how much? Would it lower any probabilities (other than the obvious complementary ones)?

More complex questions can be addressed with changes to the model. For instance, this model is based on individual participation, neglecting the effects of collaboration by several people on a single papers. Modified models would permit questions concerning increasing the number of papers presented or the number of sessions chaired. Changes in assumptions might also be made. For instance, this model assumes that the transitions from a given state are independent of the previous state. Is this assumption valid? This model also assumes that the transition probabilities have been constant over time. Is this assumption valid?

This problem is small enough for easy explication, yet rich enough to provide a variety of challenging exercises for operations research students.

REITERATING THE CAVEATS

1. The data demonstrate that there are differences between the activity level of new people and those who have a history of contributions; however, in the early meetings in the database, the groups cannot be completely separated. This impacts estimates of accessions of new people, estimates of distributions of new people's contributions at their first meeting, and estimates of the transition probabilities at all levels. The approximations here might be improved upon.
2. The conclusions on retirement probabilities are all based on inferences. These inferences might be sharpened.
3. The transition tables for the states are based on the entire database; however, the problems of first-timers who are not new people and of retirement could be removed by removing the first five meetings and the last six meetings from the analysis. Unfortunately, that leaves only 10 meetings in this database and consequently reduces the absolute number of transitions.
4. Probabilities of state transitions do not directly address relative importance of the source states for a given state. No single model is likely to answer all interesting questions.

5. Data problems could be reduced by a larger volume of data. In this problem, that means either a longer sequence of meetings data or a larger population.

CONCLUSIONS

The data show that the original intuitive concept about session chairs was correct: session chairs are very likely to chair sessions at future meetings and presenters are more likely to chair sessions as their number of presentations increase. A score of four (four presentations) or more is a good cut-off strategy. The data also show that only 10% of the people contribute again after missing six consecutive meetings and that the 10% is spread over a wide range of gap sizes. Thus a hiatus of six meetings provides a useful means of reducing appeals for session chairs by reason of probable retirement.

Some information, beyond the direct questions posed earlier, may be deduced from the data. The states represented by ovals (except state X) in Figure 5 have probabilities of making no further contributions of from 17% to 28%; whereas the states represented by rectangles have a 7% probability of retiring. This model identifies the people in these latter states as the MAS core contributors. Further, the fact that 67% of the contributions at each meeting result from new people indicates the value of inexpensive mass appeals in addition to personal appeals to high probability sources.

The model developed here is rough, but good enough for the questions demanded of it. The details might change; but clearly the model is applicable to any category of professional meeting where repeat contributions are a significant part of the presentations. Beyond answering questions about MAS contributors, the exercise of creating this model may have educational value because it shows some of the problems in extracting the necessary information from extant data and the process of creating a Markov model. It also provides a set of interesting classroom exercises. Further, Markov models appear to be used less frequently than some other classes of models. This exposition may aid some practitioner who needs a different way of looking at his or her problem.

REFERENCES

Hartley, Dean S., III. "Military Operations Research: Presentations at ORSA/TIMS Meetings," *Operations Research*, Vol 40, Number 4, July-August 1992.

Hartley, Dean S., III. "MAS CHAIRMAN: Honoring Contributors to MAS," *Phalanx*, October 1994.

Hillier, Frederick S. and Gerald J. Lieberman. *Introduction to Operations Research*, McGraw-Hill Publishing Company, New York, 1990.

ENDNOTE

[1] The author would like to thank Dr. Sam Parry of the Naval Postgraduate School for some of the concepts incorporated in this section.

RIST PRIZE CALL FOR PAPERS

MORS offers two prizes for best papers—the Barchi Prize and the Rist Prize. The Rist Prize will be awarded to the best paper in military operations research submitted in response to this Call for Papers. The Barchi Prize will be awarded to the best paper from the entire 65th symposium, including Working Groups, Composite Groups, and General Sessions.

David Rist Prize: Papers submitted in response to this call will be eligible for consideration for the **Rist Prize**. The committee will select the prize-winning paper from those submitted and award the prize at the 66th MORSS. If selected, the author(s) will be invited to present the paper at the 66th MORSS and to prepare it for publication in the MORS journal, *Military Operations Research*. The cash prize is $1000. To be considered, the paper must be mailed to the MORS office and postmarked no later than **September 30th, 1997.** Please send the original, three copies and the disk.

Richard H. Barchi Prize: Author(s) of those papers selected as the best from their respective Working Group or Composite Group, and those of the General Sessions at the 65th MORSS will be invited to submit the paper for consideration for the **Barchi Prize**. The committee will select the prize-winning paper from among those presented, nominated and submitted. The prize will be presented at the 66th MORSS. The cash prize is $1000. To be considered, the paper must be mailed to the MORS office and postmarked no later than November 28, 1997. Please send the original, three copies and a disk.

Prize Criteria

The criteria for selection for both prizes are valuable guidelines for presentation and/or submission of any MORS paper. To be eligible for either award, a paper must, at a minimum:

- Be original and a self-contained contribution to systems analysis or operations research;
- Demonstrate an application of analysis or methodology, either actual or prospective;
- Prove recognizable new insight into the problem or its solution; and
- Not previously been awarded either the Rist Prize or the Barchi Prize (the same paper may compete for but cannot win both prizes.)

Eligible papers are judged according to the following criteria:

Professional Quality

- Problem definition
- Citation of related work
- Description of approach
- Statement of assumptions
- Explanation of methodology
- Analysis of data and sources
- Sensitivity of analyses (where appropriate)
- Logical development of analysis and conclusions
- Summary of presentation and results

Contribution to Military Operations Research

- Importance of problem
- Contribution to insight or solution of the problem
- Power of generality of the result
- Originality and innovation

Statistical Validation of a Communications Network Simulation

Ann E. M. Brodeen and Malcolm S. Taylor
U.S. Army Research Laboratory

ABSTRACT

Simulation is a widely accepted means of analyzing systems that are too complex to model analytically. Most communications systems fall into this category. But when a continuing program of verification and validation is not maintained, the credibility of the simulation suffers and the value of analyses that the simulation supports is diminished. The primary goal of any verification and validation process is to enhance both the correctness of a simulation and the confidence placed in its results. A persistent challenge facing the modeler is to develop a process that is both feasible and compatible with an organization's needs, and that is widely applicable.

Multivariate statistical procedures can be used to assess the agreement between simulated predictions and empirical observations. This paper describes such a test that is useful for the validation of simulations of battlefield communications networks. The procedure is applied to a simulation that was developed to duplicate an experimental configuration in which messages were passed over a communications network using the combination of the Tactical Fire (TACFIRE) Direction System protocol and Single Channel Ground and Airborne Radio System (SINCGARS) Combat Net Radios (CNR).

LIMITED BANDWIDTH TACTICAL NETWORKS

The purpose of a communications network is to serve as a carrier of information from one location to another. The effective distribution of information can improve the decision process on the battlefield, whereas the consequences of making decisions based on obsolete information can be catastrophic. The maximum available bandwidth of the VHF-FM radios still utilized by lower echelon units is only 1,200 bits per second, a very limited data exchange rate. On a limited bandwidth tactical network, the number of nodes and the amount of information to pass can be large, especially during peak battle periods.

To measure a network's effectiveness, a determination must be made of whether the messages arrive at their destination intact and in time to be useful. The amount of correctly passed information is referred to as "network throughput," and the amount of time required to pass that information as "network delay." There are a number of conditions that can impact throughput and delay, including the number of messages to be transmitted, the size of the messages, the number of nodes on the network, the communications protocol, and the communications hardware. If the effects of these factors on network performance are better understood, attempts to optimize the network's effectiveness are more likely to succeed.

One way to examine the interactions of network parameters is through simulation. Simulation is a widely accepted means of analyzing real-world systems that are too complex to model analytically. Most communications networks fall into this category. The simulations commonly require as input the probability that two or more messages will collide, the expected delay in message transmission, or the arrival rate of messages at a given node, and then extrapolate those estimates to a complex network of multiple nodes. This approach is usually taken to simplify the simulation but requires stringent assumptions that may result in an unrealistic representation of the protocol.

A computer simulation is only a surrogate for actual experimentation with an existing or conceptual system. Simulation credibility suffers and the value of analyses the simulation supports is reduced when a program of continuing verification and validation is not undertaken. A fundamental goal of validation is to ensure that a simulation is developed that can be used by a decision-maker to arrive at the same decision that would have been made if it were possible to experiment with the actual system. Validation should serve to increase both the logical correctness of a simulation and the confidence placed in its results. The challenge confronting modelers is to develop a validation process that is both feasible and effective and sufficiently general to allow its application to a broad class of simulations. It is not uncommon to find several groups in a military organization each developing a network simulation that performs essentially the same tasks; the differences usually lie in the assumptions and/or definitions of simulation responses. Ideally, a validation procedure should be able to accommodate the simultaneous comparison of several candidate simulations.

Application Areas: C3

OR Methodologies: Nonparametric Statistics

This initiative is part of a broader-based Army research program whose goal is to improve the ability of communications networks to deliver critical information on the battlefield when and where it is needed despite a rapidly changing and often hostile environment. It will also support the ongoing effort to formalize the validation process for communications network simulations that, in turn, provide the groundwork for testing hypotheses throughout the research program. This formalization needs to be readily transmitted to other organizations that rely on communications network simulations for their analyses.

A TACTICAL NETWORK EXPERIMENT

A controlled laboratory experiment was conducted at the U.S. Army Research Laboratory's (ARL) Command, Control, Communications, and Computers (C4) Research Facility to quantify the effects of message arrival rate and message length on the throughput and delay of a small combat radio network using TACFIRE protocol over SINCGARS radio channels [1]. Measurements were taken on a number of network parameters, including network utilization, a measure of time a network is occupied with message transmissions. Network throughput and delay, along with utilization, will be important components of the validation procedure to be described in the next section.

The experimental setup consisted of four nodes, each of which was a SUN workstation, communicating over a combat radio network. Each node contained a message driver providing communications loading, and data collection software to log the sending and receipt of messages and acknowledgments, as well as information on queues. The nodes were connected to modems to enable communications via radios using a specified tactical net-sensing algorithm and communications protocol. To minimize error rates, which act as an obscurant to the parameters of interest, the radios were placed no more than 3 ft apart and were, therefore, set to low power. Resistor loads were used in place of antennas to avoid interference. Figure 1 illustrates the experimental configuration.

A scenario generator was written to create "messages" of character strings of a specified length and arrival rate over a 1-hr period. Four message arrival rates emulated the rate of actual user-generated messages and specific nodes' ability to respond to incoming messages. In the experiment, the number of messages generated and queued (but not necessarily transmitted) for transmission each hour by each node was assumed to be a mutually independent Poisson-distributed random variable, a common assumption in communications simulation [2]. The messages were equally distributed among the four nodes. For example, if the arrival rate was 2,000 messages/hour, the scenario generator created a file of 500 messages for each node. A message was assumed to enter network service when it reached the modem. Once the message was generated, the communications protocol added several layers of information to ensure that the message arrived at its destination. This included five error correction/detection bits for each seven-bit character, four synchronization characters, and a preamble to bring the transmitter to full power before the message was sent. Acknowledgments, though shorter in length, were wrapped with similar overhead bits.

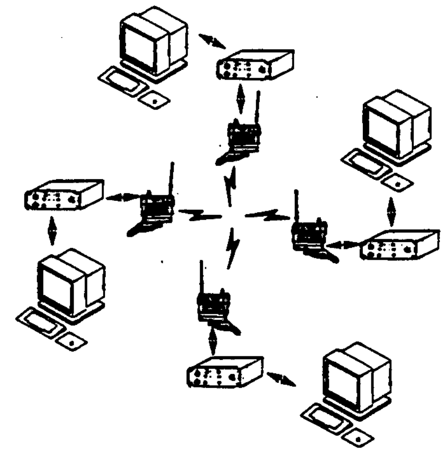


Figure 1. Hardware configuration for the laboratory experiment.

Four levels of message arrival rate were tested with each of 4 levels of message length, yielding 16 test combinations. The levels of interest for message arrival rate were 100, 250, 350, and 500 messages per node. The levels of interest for message length were 48, 144, 256, and 352 characters.

It was determined that the shortest reasonable time to test any 1 of the 16 combinations was 1 hr. This meant that a minimum of 16 hr was required for a single experimental replicate, which realistically could not be completed in 1 day. A data collection scheme known as a randomized incomplete block design (see e.g., Montgomery [3]) was constructed in order that day-to-day variability would not influence the resulting data analysis. The assignment of the combinations into blocks was based on a scheme that ensured that the effects of message arrival rate and message length, as well as their interaction, on network throughput and network delay could be accurately measured. The experiment was replicated three times to ensure the incomplete block design was balanced, facilitating analysis of the data and increasing the confidence placed in conclusions to be drawn.

A STATISTICAL PROCEDURE FOR SIMULATION VALIDATION

In assessing the fidelity of a computer simulation of a real-world communications network system, it is important to effect the comparison of the system attributes simultaneously (i.e., the corresponding measurements of network throughput, network delay, and any other network parameters chosen for study should be considered in aggregate, since these measurements are not independent of each other). The discussion of how this can be accomplished is facilitated by the introduction of some notation.

Let

$$X_j^k = \left(X_{1j}^k, X_{2j}^k, \ldots, X_{pj}^k\right)' \tag{1}$$

represent the vector of measurements taken on an arbitrary system. For our immediate purpose, the parameter p will be equal to 3, since we will consider the triple of network attributes (throughput, delay, utilization). The index k distinguishes real-world and simulated measurements (e.g., $k = 1$ denotes real-world, and $k = 2$ denotes simulated). The index j counts the number of observations taken, which may differ between the real-world and simulated systems. In general, p is an arbitrary integer, $k = 1, 2, \ldots, c$, and $j = 1, 2, \ldots, n_k$. The general notation will be suppressed in what follows for clarity of presentation, but it should be strongly emphasized that the method to be described is applicable in far more general situations than the application detailed here. Specifically, more than 3 parameters ($p > 3$) may be used and more than 1 simulation ($k > 2$) may be compared simultaneously.

The basic idea is as follows. To compare the real-world observations X_j^1, $j = 1, 2, \ldots, n_1$, and simulated observations X_j^2, $j = 1, 2, \ldots, n_2$, for agreement, one might equally ask if the two sets of observations appear to have come from the same population. If the answer is yes, the two sets are difficult to distinguish between, and acceptance of the simulation as a faithful emulation of the real-world may be appropriate. If the answer is no, the simulated and real-world data appear different, and the validity of the simulation is called into question.

To proceed with this approach the data is organized into a matrix

$$\mathbf{X} = \begin{bmatrix} X_{11}^1 & .. & X_{1n_1}^1 & .. & X_{1n_c}^c \\ \cdot & & \cdot & & \cdot \\ \cdot & & \cdot & & \cdot \\ \cdot & & \cdot & & \cdot \\ X_{p1}^1 & .. & X_{pn_1}^1 & .. & X_{pn_c}^c \end{bmatrix} \tag{2}$$

which is simply the $p \times \mathrm{N}$ $(= \sum n_k)$ matrix whose columns are formed from the vectors X_j^k defined in Eqn. 1; that is,

$$\mathbf{X} = \left(X_1^1, \ldots, X_{n_1}^1; X_1^2, \ldots, X_{n_2}^2\right) \tag{3}$$

The next step is to transform the entries in the matrix $\mathbf{X} = (X_{ij}^k)$ as follows. Within each row i of the matrix $\mathbf{X}$, order the entries from smallest to largest:

$$X_{i(1)}^k \leq X_{i(2)}^k \leq \cdots \leq X_{i(\mathrm{N})}^k \tag{4}$$

and assign to the smallest value, rank 1; the second smallest value, rank 2; ...; the largest value, rank $\mathbf{N}$. Replacing (or transforming) the entries

in the matrix **X** by their corresponding ranks gives rise to a matrix

$$\mathbf{R} = \begin{bmatrix} R_{11}^{1} & .. & R_{1n_1}^{1} & .. & R_{1n_c}^{c} \\ \cdot & & \cdot & & \cdot \\ \cdot & & \cdot & & \cdot \\ \cdot & & \cdot & & \cdot \\ R_{p1}^{1} & .. & R_{pn_1}^{1} & .. & R_{pn_c}^{c} \end{bmatrix}. \quad (5)$$

If the data come from a common population, then in each row of the matrix **R** the assignment of ranks should be random. If a systematic assignment of ranks seems to be occurring—especially if smaller ranks are associated with one set $\{X_{ij}^{k}\}$ a and larger ranks with another $\{X_{ij}^{k'}\}$ —the inference that the data are not from a common population may be justified.

The comparison of the data sets is made as follows. For each row i, compute the mean of the ranks assigned to the k^{th} sample

$$T_i^k = \frac{1}{n_k} \sum_{j=1}^{n_k} R_{ij}^k,\ i = 1, 2, 3 \text{ and } k = 1, 2.$$

If the samples are from a common population, each row of the matrix **R** is a random permutation of the integers 1, 2, ..., N, and the sample means T_i^k should be close in value to the overall mean $E_i = (N + 1)/2$ (the mean of the integers 1, 2, ..., N). A test for agreement between data sets based on this observation can then be constructed by forming the contrasts

$$T_i^k - E_i,\ i = 1, 2, 3 \text{ and } k = 1, 2, \quad (7)$$

each of which is expected to be numerically small if agreement between simulated and real-world data is good. An expression involving all the contrasts that will be sensitive to the numerical largeness of any contrast seems to be an appropriate statistic for a simultaneous comparison. One function that accommodates this goal is suggested by Puri and Sen [4], and in this example takes the form

$$L_N = \sum_{k=1}^{2} n_k\ [(T^k - E)\ V^{-1}(\mathbf{R})\ (T^k - E)'],$$

where

$$T^k = \left(T_1^k, \ldots, T_3^k\right) \text{ and } E = \left(E_1, \ldots, E_3\right). \quad (8)$$

(Puri and Sen's expression is more general and involves summation over $k = 1,2, ..., c$.) L_N is a weighted sum of a quadratic form in $(T^k - E)$, and $V^{-1}(\mathbf{R})$ is the inverse of the covariance matrix of **R**. The quadratic form L_N is mathematically attractive because the correlation structure between the variates $i = 1, 2, ..., p$ is taken into account through the covariance matrix $V(\mathbf{R})$ [5]. Scaling of the variates, a necessary precaution to ensure that an artificial dominance of one variate over another due simply to scale of measurement does not occur, was automatically accomplished by replacing the original measurements with their rank assignments.

The comparison of the three-dimensional data sets is now reduced to expression by a single number—the statistic L_N. Large values of L_N reflect a disagreement between simulated and real-world data; small values of L_N suggest agreement. What constitutes "large" or "small" values of L_N remains to be resolved, but can be accomplished in an elegant and straightforward manner following a procedure introduced by Fisher [6].

The matrix **R** can have in general $(\mathbf{N}!)^p$ different realizations. The statistic L_N can be evaluated for each of these realizations and the (not necessarily distinct) values ordered from smallest to largest. By observing where the value of L_N computed for the specific experimental/simulation data combination under analysis (say, $L_N = L_N^*$) falls in this ranking, an assertion of how unusual that value is may be made. If a small fraction, say 5%, of potential L_N values equal or exceed the observed value L_N^*, then an unusually large value of L_N^* has occurred, and a disparity between simulated and real-world data may be assumed. Otherwise, such a distinction cannot be made and the simulation may be regarded as valid.

The methodology detailed in this section parallels the traditional statistical hypothesis test framework. Specifically, it is a multivariate (there are three variates) nonparametric (no distribution assumptions are made) rank (the data were

transformed into ranks) test. The procedure attributed to Fisher is known as a randomization or permutation method, depending on some specifics of data collection. An important consideration is that the statistic L_N defined in Eqn. 8 may be replaced by any other expression that the analyst deems appropriate, an unusual option that adds significantly to the power of the methodology.

EXAMPLE AND DATA ANALYSIS

We are considering here the special case of comparing two systems, real-world and simulated, on the basis of several carefully selected performance measures. Although data for a number of measures of performance were collected during the laboratory experiment previously described, comparisons between experimental and simulation results will be limited to the continuous random variables—network throughput, network delay, and network utilization. Output from a simulation built utilizing the tools of a commercially available software package dedicated to communications network modeling was compared against the results from the laboratory experiment. Insofar as possible, initial conditions between the laboratory experiment and the simulation were matched. The stochastic simulation was run seven times, providing seven simulation replications to compare with the three laboratory replicates.

Network throughput is calculated as the average number of information bits that were successfully transmitted and acknowledged over a 1-hr test cell. Throughput does not include such overhead as the acknowledgments themselves, or, in the event of collisions, message retransmissions. It does, however, include error detection/correction bits and synchronization characters. Network delay is the average time that passes between a message's arrival at a host's modem until the acknowledgment returns to the host. Messages that were never completely serviced during the running of a test cell were not considered in computing network delay. Network utilization for a particular time interval is the amount of time spent actually transmitting messages, message retransmissions, or acknowledgments during the interval, divided by the amount of time in the interval. Messages, retransmissions, and acknowledgments include a preamble and other protocol overhead in addition to actual transmission bits.

Although 16 combinations of message arrival rate and message length were included in the laboratory experiment, only 8 were chosen for validation purposes. The eight combinations were chosen selectively (e.g., it was important to evaluate the simulation at the two extremes of both parameter ranges [i.e., arrival rate of 400 messages and message length of 48 characters; arrival rate of 2,000 messages and message length of 352 characters]). The compatibility between experimental and simulated data was evaluated separately for each combination of arrival rate and message length. No attempt to create an omnibus test was undertaken. To have done so would incur an attendant loss of information specific to any particular pairing, which was undesirable at this stage of validation.

The matrix of ranks $\mathbf{R}$ for the combination (2,000 messages, 144 characters) is shown in Figure 2. There are three vectors of real-world measurements and seven vectors of simulated measurements which form the ten columns of $\mathbf{R}$.

$$\mathbf{R} = \begin{bmatrix} 3 & 2 & 1 & 4 & 6 & 7 & 8 & 10 & 9 & 5 \\ 6 & 9 & 7 & 10 & 1 & 8 & 3 & 2 & 4 & 5 \\ 1 & 2 & 3 & 4 & 6 & 7 & 9 & 10 & 8 & 5 \end{bmatrix}$$

Figure 2. Rank matrix for the combination (2,000 messages, 144 characters).

The number of permutations of the ranks is 10!, and the L_N statistic needs to be calculated $10!/3!7! = 120$ times. Since this validation study deals with small values of $\mathbf{N}$ (= 10) and p (= 3), the L_N statistic may be easily evaluated for all the permutations of the ranks. In the actual computation, the ranks were multiplied by $1/(\mathbf{N} + 1) = 1/11$. This constant multiplier does not affect the outcome (and, as such, is not really necessary) but it causes the statistic L_N to reduce in the univariate case $p = 1$ to a well-known expression in nonparametric statistics known as the Kruskal-Wallis test, and, as such, holds mathematical appeal.

The statistic L_N given in Eqn. 8 was evaluated for the data in the rank matrix shown in Figure 2 and determined to be $L_N^* = 6.73$. The sig-

nificance (or p-value) associated with L_N^* is the proportion of the 120 values in the reference set that equal or exceed L_N^*. In this instance, the number of values was determined to be nine, which translates into a significance level of 0.075.

The same procedure was applied to the eight test combinations selected for validation of the simulation. The observed values of the test statistic and the corresponding p-values are summarized in Table 1. Note that in five of the eight combinations tested, the simulation results did not agree with the experimental data. The level of (dis)agreement is impartially quantified by the statistical procedure. These results do not mean that the simulation should be abandoned, but some revision is clearly implied. This approach is consistent with a program of continuing improvement in simulation development mentioned at the onset and reinforces the decision to not seek an omnibus test at the onset of simulation validation.

Input Values (msgs, chars)	L_N^*	p-value	Outcome
400, 48	7.18	0.04	Reject
400, 256	9.03	0.01	Reject
1,000, 144	6.97	0.06	Accept
1,000, 352	9.65	0.01	Reject
1,400, 48	7.83	0.02	Reject
1,400, 256	6.58	0.10	Accept
2,000, 144	6.73	0.08	Accept
2,000, 352	9.21	0.01	Reject

Table 1. Validation Results

The testing procedure, and the conclusion reached regarding the validity of a simulation, depends to some extent on the choice of statistic used for comparison. Theoretical considerations of the communications network had determined *a priori* that a correlation structure exists among the variates, and so successive application of commonly used univariate procedures was inappropriate [7].

An alternative approach to the problem described in this paper (and one suggested by a referee) is a more classical statistical procedure known as multiple analysis of variance (MANOVA). MANOVA is developed under strong distribution assumptions. The response vector—here the triple (throughput, delay, utilization)—is assumed to follow a multivariate normal distribution, and the covariance structure of the variates is assumed constant across the conditions under experimental control—here (data source, message length, arrival rate).

After a preliminary screening of these particular data, and lacking historical data or theoretical impetus, we were uncomfortable with the inherent assumptions and chose instead a nonparametric approach [8,9]. However, there is an argument to be made related to robustness of the MANOVA procedure and its value as a screening device. A complete data analysis almost always benefits from both parametric and nonparametric approaches and, along with the liberal use of graphics, should be included in the initial stages of every careful statistical inquiry. MANOVA procedures with strong graphic support are available in most of the popular statistical packages, which further encourage their use. The nonparametric procedure detailed here, while straightforward and easily programmed, is not as likely found in off-the-shelf software.

SUMMARY

As reliance upon computer simulations to model processes that resist analytical description and to support decision making increases, so does the need to validate the simulations. An impartial approach to simulation validation may be taken through statistical hypothesis testing. An application of a nonparametric multivariate statistical procedure to assess the validity of a communications network simulation was detailed. The method discussed offers considerable flexibility to the analyst charged with maintaining the fidelity of the simulation effort and holds the promise of application in many more general situations.

REFERENCES

[1] Kaste, V. A. T., A. E. M. Brodeen, and B. D. Broome. "An Experiment to Examine Protocol Performance Over Combat Net Radios." BRL-MR-3978. U.S. Army Research Laboratory, Aberdeen Proving Ground, MD, June 1992.

[2] Law, A. M., and W. D. Kelton. Simulation Modeling and Analysis. New York: McGraw-Hill, Inc., 1991.

[3] Montgomery, D. C. Design and Analysis of Experiments. New York: John Wiley & Sons, 1991.

[4] Puri, M. L., and P. K. Sen. Nonparametric Methods in Multivariate Analysis. Malabar, FL: Krieger Publishing Company, 1993.

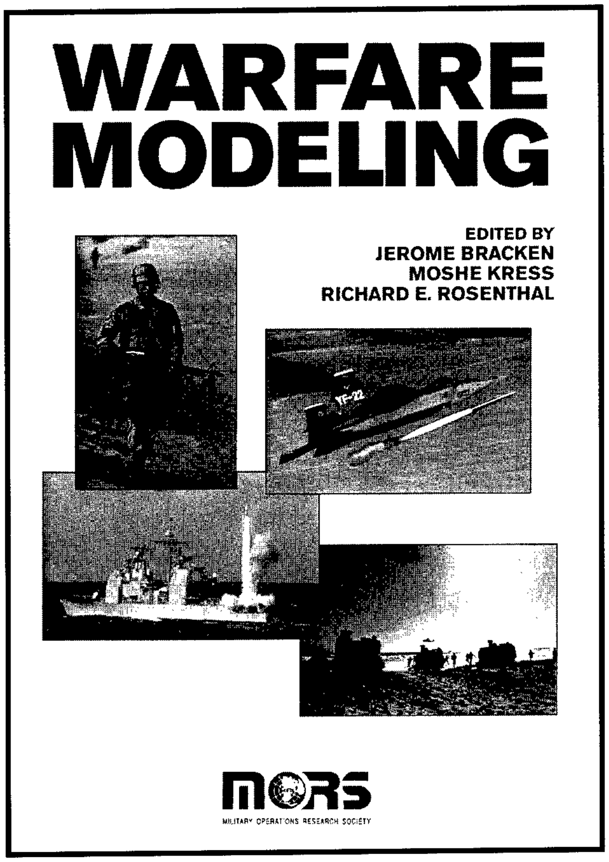
[5] Good, P. Permutation Tests. New York: Springer-Verlag, Inc., 1994.

[6] Fisher, R. A. Design of Experiments. Edinburgh: Oliver and Boyd, 1935.

[7] Green, P. E., Jr. Computer Network Architectures and Protocols. New York: Plenum Press, 1982.

[8] Black, U. D. Data Communications, Networks, and Distributed Processing. Reston, VA: Reston Publishing Company, Inc., 1983.

[9] Bertsekas, D. Data Networks, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1992.



ORDER FORM

Warfare Modeling

Published by the
MILITARY OPERATIONS RESEARCH SOCIETY
in cooperation with John Wiley & Sons
600 pages, $35.

The Military Operations Research Society (MORS) serves the students, developers, practitioners and clients of military operations research through the ongoing production of high-quality, low-cost meetings, books and periodicals.

Warfare Modeling was selected for publication by MORS because it contains 25 state-of-the-art chapters, addressing a wide class of operations research models that are of central importance in military planning, analysis and operations. Research on this subject matter seldom appears in the open literature. Most of it appears in reports and documents of government agencies and advisory corporations. This book is an invaluable reference for military OR professionals as well as for a more general audience of researchers and practitioners. The scope of models included—analytic and simulative, stochastic and deterministic, simple and complex, applied and theoretical, domestic and international—is very broad.

Warfare Modeling was edited by Jerome Bracken of Yale University, Moshe Kress of Israel's Center for Military Analyses, and Richard E. Rosenthal of the US Naval Postgraduate School. The Foreword is by Wayne P. Hughes, Jr., a retired Navy captain and MORS Fellow.

The book is comprised of seven sections:

(1) Theater-Level Modeling
(2) Mathematical Models of Combat
(3) Historical Analysis
(4) Weapon System Analysis
(5) Command, Control, Communications and Intelligence
(6) Cost Effectiveness Analysis
(7) Modeling in Support of Operations Desert Shield and Desert Storm

Send orders to: Military Operations Research Society (MORS)
101 S. Whiting Street, Suite 202, Alexandria, VA 22304
703-751-7290; FAX 703-751-8171; e-mail morsoffice@aol.com

Please send ______ copies of *Warfare Modeling* . . . @$35.00 each __________
Freight: U.S. . . . . @ $5.00 each __________
Canada . . . . @ $8.00 each __________
Other International . . . . @$18.00 each __________
Virginia residents add sales tax . . . . @ $1.58 each __________
. . . . .TOTAL __________

Name: ____

Address: ____

City: ________________________ State: ____________ Zip: __________

Payment must be made in US dollars drawn on a US bank.

If paying by credit card

❑ Visa
❑ Master Card
❑ American Express

____
Credit Card Number

____
Expiration Date

____
Signature

A Data Analysis of Success in OCS, The Use of ASVAB Waivers, and Race

R.R. Read
L.R. Whitaker
Naval Postgraduate School

ABSTRACT

Applicants for Officers Candidate School (OCS) can receive a mental aptitude qualification waiver based upon their scores on the electronics portion of the Armed Services Vocational Aptitude Battery (ASVAB). The question arises whether the candidates that receive a waiver have the same success rate in OCS as those who do not. From OCS records there is strong evidence that the overall rate of success in OCS is smaller for those candidates who hold a waiver than for those candidates who do not hold a waiver. However, closer inspection of the data reveals that success rates change with race in such a way that, for each racial group, the presence or absence of a waiver is not noticeable. That is, success is conditionally independent of waiver. This independence is lost when the conditioning is removed. Thus what initially seemed to be a waiver policy issue is confounded by the rate of granting waivers by race and differences in success rates by race. The OCS data are studied to expose this conundrum and to develop sharper models for success in OCS.

I. INTRODUCTION

The accession of officers into the Marine Corps via OCS includes the use of one of three mental aptitude test scores: Armed Services Vocational Aptitude Battery Electronics Repair Composite (ASVAB), the Scholastic Aptitude Test (SAT), and the American College Test (ACT). Historically, 55% of the officers entering use the first of these three, and the qualification threshold is a score of 120. But a candidate can receive a waiver of this minimum provided his score is 115 or better. This analysis treats only those using the ASVAB test.

Based on data collected over the fiscal years 1988 through 1992 and broken out by race, personnel at the Manpower Analysis (MA) Branch at Marine Corps Headquarters noticed that success at the Officer Candidate School (OCS) appears to be independent of whether an officer has received an ASVAB waiver. Specifically, there are four recorded racial groups, Caucasian, Black, Hispanic, and Other. The Other group consists largely of American Indian, Alaskan Native, Asian, and Pacific Islander. When collapsed over time, the four 2 x 2 contingency table tests for independence yield the chi square test statistics .6678, 2.841, .7983, .5767 for the respective races, each with one degree of freedom. None of these are significant. However, when the data are further collapsed over race and a single test for independence is performed, then the relationship is highly significant. This latter 2 x 2 table appears in Table 1. The chi square statistic is 11.87 and the p-value is 0.00057.

On the surface, it appears that we have contradictory results. On the one hand, OCS candidate success and the presence of a waiver are independent when Caucasians, Blacks, Hispanics and Others are considered separately. On the other hand, there is dependence in the collapsed table when race is not accounted for, with strong evidence that the chance of success without a waiver (76%) is greater than that with a waiver (72%).

	Waiver	No Waiver	Total
Success	754	7449	8203
Failure	299	2303	2602
Total	1053	9752	10805

Table 1. Macro Analysis of Success and Waiver

A short answer to the contradiction can be obtained through an interpretation of the two success rates. They are not significantly different for waiver and non-waiver within racial groups. But the rates change sharply from group to group. Indeed, the use of the waiver varies markedly from group to group and, to a lesser extent, from year to year. This is surely related to the implementation of the Marine Corps Affirmative Action Plan.

This paper contains an explanation of the contradiction and attention is drawn to other interesting facets as well. In Section II the raw data are presented and all 2 x 2 tables of success/failure by waiver/non-waiver are studied for each year/racial group pair. Generally, independence is tenable. To explain the non-independence, the full data, aggregated over years and with race as a factor, are then subjected to a log-linear analysis in Section III. In Section IV, we fit models with time as a factor including the use of the waiver by year and race. These models could be valuable because an ill-advised long-term overuse of the waiver could lead to inequities in the future advancement to higher rank [3].

Application Areas:
Manpower and Personnel, Training

OR Methodologies:
Statistics, Categorical Data Analysis

Categorical data is prevalent in military OR. Thus, we take a careful look at the data and provide details that would normally be omitted so that certain usage may be illustrated. In particular, in the next section, attention is drawn to the rather interesting effects when conditional tests are used, and in Section III the steps for fitting a loglinear model are presented.

The factors of interest are success or failure of candidates in the OCS program, whether the candidate used an ASVAB (lower mental category) waiver, fiscal year, and race. The data (see Table 2) consists of counts

$$D_{ijkl}$$

where $i = 1, 2$ indicates success or failure, $j = 1, 2$ indicates presence or absence of waivers, $k = 1, \ldots, 5$ indicates the fiscal year FY88 to FY92 and $l = 1, \ldots, 4$ indicates race, in the order given earlier.

Candidates Qualifying with ASVAB Waiver

	FY	White	Black	Hispanic	Other	Total
Success in OCS	FY88	100	11	10	12	133
	FY89	142	37	12	20	211
	FY90	102	30	20	11	163
	FY91	77	22	14	2	115
	FY92	70	36	22	4	132
	Total	491	136	78	49	754

	FY	White	Black	Hispanic	Other	Total
Failure in OCS	FY88	22	8	5	1	36
	FY89	30	15	11	7	63
	FY90	35	16	10	3	64
	FY91	21	22	6	3	52
	FY92	45	31	8	0	84
	Total	153	92	40	14	299

Candidates Qualifying without ASVAB Waiver

	FY	White	Black	Hispanic	Other	Total
Success in OCS	FY88	1113	48	48	95	1304
	FY89	1533	56	80	111	1780
	FY90	1263	77	76	109	1525
	FY91	1013	58	78	39	1188
	FY92	1390	87	108	67	1652
	Total	6312	326	390	421	7449

	FY	White	Black	Hispanic	Other	Total
Failure in OCS	FY88	234	14	16	31	295
	FY89	323	18	22	35	398
	FY90	350	50	41	38	479
	FY91	430	35	38	24	527
	FY92	481	50	48	25	604
	Total	1818	167	165	153	2303

Table 2. Frequency Counts by Category

II. INDIVIDUAL CONTINGENCY TABLES

Suppose the full data are broken into twenty (5 years, 4 races) 2 x 2 contingency tables and subjected to individual analyses. It is instructive to apply the most often used procedures to each and gain experience in their use and effect.

Let us simplify the notation and let $n_{ij} = D_{ijkl}$ be the counts with year and race held fixed, $i = 1, 2$ indicates success or failure in OCS, and $j = 1, 2$ indicates presence or absence of waiver, respectively. Under independence the expected frequencies are estimated by

$$\hat{m}_{ij} = n_{i+}n_{+j} / N \qquad \text{with } N = \sum\sum n_{ij},$$

and the plus indicates summation over the replaced subscript. The estimated frequencies under independence based on Table 1 are given in Table 3. The estimated success rate is 76% in both instances. The familiar Pearson Chi Square and Log Likelihood statistics are given by

$$\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{2}\left(n_{ij} - \hat{m}_{ij}\right)^2 / \hat{m}_{ij}$$

$$G^2 = 2\sum_{i=1}^{2}\sum_{j=1}^{2} n_{ij} \ln\left(n_{ij} / \hat{m}_{ij}\right).$$

Each is asymptotically distributed as chi square with one degree of freedom.

	Waiver	No Waiver	Total
Success	799	7404	8203
Failure	254	2348	2602
Total	1053	9752	10805

Table 3. Estimate Frequencies under Independence

The use of the odds ratio is also popular especially in 2 x 2 tables. It summarizes the strength and type of dependence between the two categories. Letting $\{\Pi_{ij}\}$ be the cell probabilities, the odds ratio is defined by

$$\theta = \Pi_{11}\Pi_{22} / \Pi_{12}\Pi_{21}$$

and, in our context, represents the odds of OCS success using waivers divided by the odds of success without the use of waivers. The null value $\theta = 1$ represents "no effect" of waivers, or independence. The maximum likelihood estimator of θ is

$$\hat{\theta} = n_{11}n_{22} / n_{12}n_{21}.$$

The null distribution of $\ln(\hat{\theta})$ is well approximated by the normal distribution [1]with the variance estimated by

$$\left[\hat{\sigma}(\ln \hat{\theta})\right]^2 = \sum_1^2 \sum_1^2 1 / n_{ij}.$$

Thus, a third test statistic is

$$Z = \ln(\hat{\theta}) / \left[\sum \sum 1 / n_{ij}\right]^{1/2}.$$

Concern for the use of asymptotics has led the authors to consider Fisher's Exact Test as well, [1, p60ff]. Under the null hypothesis of independence, an exact distribution that is free of any unknown parameters results from conditioning on the totals in both margins. The result is a hypergeometric distribution

$$\binom{n_{+1}}{n_{11}}\binom{n_{+2}}{n_{12}} \Big/ \binom{N}{n_{1+}}.$$

Since the totals in the margins are given, only n_{11} need be considered as variable. Its range is

$$\max(0, n_{+1} + n_{1+} - N) \le n_{11} \le \min(n_{+1}, n_{1+}).$$

Exact two-sided p-values are obtained by summing probabilities of tables that are at least as rare under the null hypothesis as the observed table. Only those tables that have hypergeometric probabilities at least as small as the observed configuration are used [2].

The results of the four procedures are given in Table 4, which contains the values of total populations, N; the odds ratios, $\hat{\theta}$; $\ln(\hat{\theta})$; the standard deviation of $\ln(\hat{\theta})$; and the four p-values. Within cells the racial levels are Caucasian, Black, Hispanic, Other, respectively. There are some blank entries for the last case because $n_{21} = 0$.

		N	$\hat{\theta}$	$\ln\hat{\theta}$	$\hat{\sigma}(\ln\hat{\theta})$	X^2	G^2	Z	Fisher
FY88	Cauc.	1469	.956	-.045	.246	.854	.854	.854	.804
	Black	81	.401	-.914	.555	.094	.104	.100	.139
	Hisp.	79	.667	-.405	.619	.511	.518	.513	.527
	Other	139	3.916	1.365	1.061	.168	.126	.198	.298
FY89	Cauc.	2028	.997	-.003	.210	.990	.990	.990	1.000
	Black	126	.793	-.232	.409	.570	.571	.570	.681
	Hisp.	125	.300	-1.204	.482	.010	.014	.012	.017
	Other	173	.901	-.104	.480	.828	.829	.828	.810
FY90	Cauc.	1750	.808	-.213	.205	.296	.304	.297	.285
	Black	173	1.218	.197	.359	.583	.582	.583	.723
	Hisp.	147	1.079	.076	.433	.861	.860	.861	1.000
	Other	161	1.278	.245	.678	.717	.712	.717	1.000
FY91	Cauc.	1541	1.556	.442	.253	.078	.070	.080	.085
	Black	137	.603	-.506	.370	.170	.172	.172	.196
	Hisp.	136	1.137	.128	.527	.808	.807	.808	1.000
	Other	68	.410	-.892	.949	.335	.342	.348	.379
FY92	Cauc.	1986	.538	-.620	.198	.002	.002	.002	.002
	Black	204	.667	-.405	.303	.180	.182	.181	.223
	Hisp.	186	1.222	.200	.448	.654	.651	.654	.828
	Other	96				.225	.116		.570

Table 4. Two-Sided p-values

Perhaps the first thing to notice is the agreement of p-values for the three asymptotic procedures. Only for the smaller values of N do they show much separation. On the other hand, the p-values for Fisher's Exact Test generally tend to be higher. The main reason for this is the conditioning on both margin totals. Such is not the case in the other procedures. By conditioning on the margin totals, the nuisance parameters are eliminated in Fisher's Exact Test while in the other three procedures they are estimated.

The differences in p-values do not lead to conflicting conclusions, however. Two cases of the twenty are significant: Hispanics '89 and Caucasians '92. In both of these cases the odds for success are smaller if waivers are used. The opposite is true for Caucasians '91, a case that might be controversial as $p \sim .08$.

III. GENERAL MODELS

The four factors; success/failure, waiver/no waiver, year (1, ..., 5), and race (1, ..., 4); are denoted as A, B, C, D, respectively. Since the total number of OCS candidates is not fixed, the data D_{ijkl} will be assumed to be generated from an independent Poisson sampling scheme, i.e., D_{ijkl} are independent Poisson random variables with respective parameters (m_{ijkl}) where $m_{ijkl} = E[D_{ijkl}]$. To interpret the results given in

the introduction we first fit a loglinear model to the counts collapsed over years, i.e., to

$$D_{ij+l} = \sum_{k=1}^{5} D_{ijkl}.$$

The saturated loglinear model parameterizes $m_{ij+l} = E\left[D_{ij+l}\right]$ as

$$\ln m_{ij+l} = \mu + \lambda_i^A + \lambda_j^B + \lambda_l^D + \lambda_{ij}^{AB} + \lambda_{il}^{AD} + \lambda_{jl}^{BD} + \lambda_{ijl}^{ABD},$$
$$i = 1,2 \qquad j = 1,2 \qquad l = 1,\ldots,4,$$

with the contrast conventions

$$\lambda_1^A = \lambda_1^B = \lambda_1^D = \lambda_{1j}^{AB} = \lambda_{i1}^{AB} = \ldots = \lambda_{ij1}^{ABD} = 0$$

and where the λ's are the effects and interaction terms corresponding to the variables A, B, D given in the superscript (e.g. λ_l^D is the effect of race and λ_{jl}^{BD} is the interaction term for waiver/no waiver and race). Using standard notation [1], this saturated model can be represented as [ABD], i.e., the third order interaction term ABD and all lower order terms made up of subsets of the variables A, B, and D are included in the model. We begin by fitting the model with all two-way interaction terms along with all main effects, i.e., the model [AB] [AD] [BD]. This gives a likelihood ratio test statistic of 2.55 with 3 degrees of freedom and a p-value of .466. This model does fit the data. To see whether a more parsimonious model can be fit we remove two-way interaction terms one at a time. This yields the model [AD] [BD]. The overall likelihood ratio test statistic is 4.84 with 4 degrees of freedom giving an acceptable p-value of .31. To see whether anything has been lost by removing the AB interaction term, we test the null hypothesis [AD] [BD] versus the alternative [AB] [AD] [BD]. The test statistic 1.99 with 1 degree of freedom has a p-value of .256. There is not enough evidence to indicate that the AB term should be included. Further, deleting terms from the [AD] [BD] model yields models with unacceptable fits, i.e., those with likelihood ratio test statistics having p-values less than .05. Finally, the standardized residuals for the [AD] [BD] model range from -.843 to 1.090. Thus, the model [AD] [BD] is selected and fits the data (collapsed over years) reasonably well.

The question now becomes, can this model account for the results that motivated the study. The probabilistic interpretation of the model [AD] [BD] is that conditional on the levels of factor D (race), the variables A and B are independent. To see this note that the joint probability mass function (pmf) of the variables A, B, C, D is

$$P_{ijkl} = \frac{m_{ijkl}}{m_{++++}},$$

for $i = 1, 2$; $j = 1, 2$; $k = 1, \ldots, 5$; and $l = 1, \ldots, 4$. The model [AB] [BD] fitted to the data collapsed over years corresponds to

$$\ln m_{ij+l} = \mu + \lambda_i^A + \lambda_j^B + \lambda_l^D + \lambda_{il}^{AD} + \lambda_{jl}^{BD}.$$

Thus the conditional pmf of A given that B is at level j and D is at level l can be found from this model to be

$$P_{i|jl} = \frac{P_{ij+l}}{P_{+j+l}},$$
$$= \frac{\exp\left\{\mu + \lambda_i^A + \lambda_l^D + \lambda_{il}^{AD}\right\}}{\sum_i \exp\left\{\mu + \lambda_i^A + \lambda_l^D + \lambda_{il}^{AD}\right\}}. \tag{3.2}$$

Since the right hand side of (3.2) is not a function of j, we see that the conditional pmf of A given B, D is the same as the conditional pmf of A given D. Thus given D, the factors A and B are independent.

However, A and B are not independent by themselves alone. The marginal probabilities of these two factors can be developed from the model (3.1) by summing

$$\exp\left\{\mu + \lambda_i^A\right\} \sum_\ell \sum_j \exp\left\{\lambda_j^B + \lambda_\ell^D + \lambda_{i\ell}^{AD} + \lambda_{j\ell}^{BD}\right\}$$

and

$$\exp\left\{\mu + \lambda_j^B\right\} \sum_\ell \sum_i \exp\left\{\lambda_i^A + \lambda_\ell^D + \lambda_{i\ell}^{AD} + \lambda_{j\ell}^{BD}\right\}$$

and forming the appropriate normalizations. The joint probability is not the product of these probabilities. Thus the model supports the observation made earlier that success of the OCS candidate is not independent of whether the ASVAB waiver has been used for entry. These

two variables are independent, however, when broken out by race.

The following probabilities help interpret the dependence between A and B. The probabilities of success given race are estimated to be .78, .64, .70, .74 for Caucasians, Blacks, Hispanics and Others, respectively. (The empirical rates and the modeled rates are the same to two decimal places.) Because success is independent of waiver status given race, these probabilities are the same for those candidates with a waiver and those candidates without a waiver. The proportions of candidates in each race which possess a waiver are .07, .32, .18, .10, and the proportions of candidates who don't possess a waiver in each race are the complementary values, .93, .68, .82, .90. The greatest proportion of candidates who don't possess a waiver are Caucasians (93%), with a good chance of success (78%). However, candidates that do utilize the waiver are divided primarily between Blacks (32%) and Hispanics (18%). Because the probability of success for these two races differ (67%) and (70%) respectively, we see that the overall probability of success with a waiver is lower than without a waiver. Also, the four success rates decrease monotonically as the four waiver use rates increase. Thus the difference in the overall success rate among those who hold a waiver and those who do not does not appear to be caused solely by the presence of waiver but by differences in success rates between races and the differences in the proportions of waivers given by race.

IV. TEMPORAL ANALYSIS

The above analysis responds to the question posed in the introduction. But it is also of interest to consider the other factor, C, the fiscal year. If including the variable race sheds light on the dependence between having a waiver and success of the OCS candidate, perhaps considering this fourth variable will add to an understanding of this data set.

Perhaps the most direct way to proceed is to consider the most general four factor model that reflects independence of factors A and B. In the notation established this would be [ACD] [BCD]. All interactions involving A and B are zero. Doing so produces a likelihood ratio p-value of .049. This is rather small for our tastes. Study of the residuals reveals two outlier cells: unsuccessful Hispanics with a waiver in FY89 and unsuccessful Caucasians with a waiver in FY92. These two cells belong to the same cases that exhibited low p-values in Table 4.

It appears that the loglinear modeling system must provide for some AB interactive terms. Accordingly we apply the strategy which fits the models with all three way and lower order terms; all two way and lower order terms; and all one way terms. Then the overall model with the fewest terms and an acceptable overall fit is used as a starting point for further deletion of terms within the chosen set. The first model fit was the one with all three way interactions. This gives an overall fit with a p-value of .0387. However, as terms are deleted the p-value increases and the model [ABC] [BCD] [ACD] gives a slightly higher p-value for overall fit of .0657. Further deletion of terms leads to the model [ABC] [BCD] [AD] with p-value .22.

The fact that the deletion of additional terms appears to improve the fit can be explained by noting the increase in the degrees of freedom. For the model with all three way interaction terms, the likelihood ratio test statistic is 21.95 with 12 degrees of freedom, deleting the ABD term increases degrees of freedom to 15 and the test statistic to 24.01 and the deletion of the ABD term increases the degrees of freedom to 19 and the test statistic to 29.548. Therefore deleting terms does not increase the test statistic very much compared to the gain in degrees of freedom.

Deleting either the ABC or BCD terms from the [AD] [ABC] [BCD] model results in models with much lower p-values for overall goodness of fit and standardized residuals that are of much larger magnitude than those of the [AD] [ABC] [BCD] model. Since the standardized residuals for this model range between -1.78 to 1.81, this model appears to give an adequate fit. In passing, we note that all AB interactive terms are modest in size.

The estimated probabilities of success given race, waiver status and fiscal year $\left(\hat{p}_{i|jkl}\right)$ are plotted against year (k) in Figures 1 and 2. There is a general decrease in the probability of success over time in all four racial groups regardless of waiver status. In fact, when the model [AD] [BD]

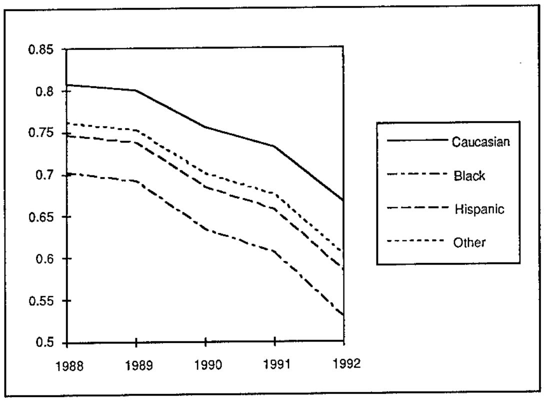


Figure 1. $\hat{P}$[S | with waiver, race, year] vs. Year

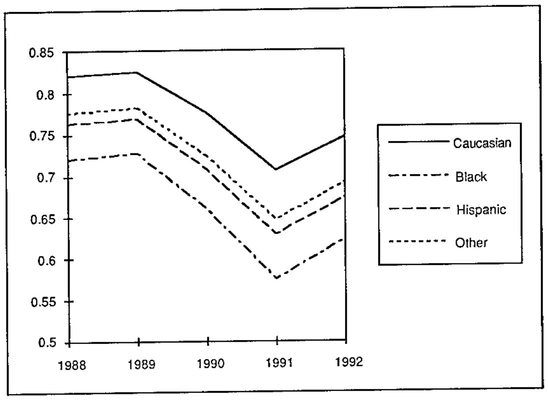


Figure 2.
$\hat{P}$[S | without waiver, race, year] vs. Year

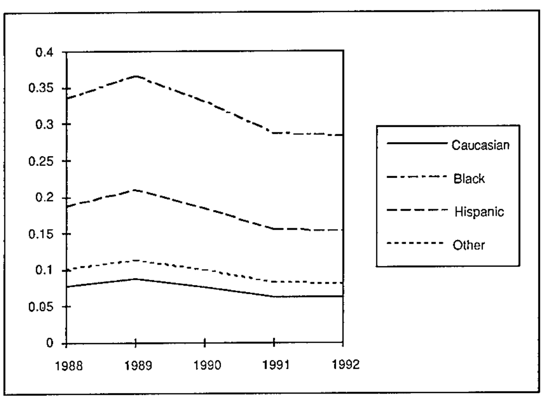


Figure 3.
$\hat{P}$[granting waiver | race, year] vs. Year

is fit to years separately, only 1992 fails to fit with a p-value = .01. It appears that for the first four years this trend is reasonably well modeled as independent of waiver status. The presence of the ABC interaction term in the temporal model is a consequence of changes in 1992, specifically the outlier cell cited earlier.

The presence of the BCD interaction term can be explained by changes in the number of waivers utilized over time. To examine this, we fit a logistic regression model where the response variable is one or zero according to whether an individual received a waiver or not, and the explanatory variables are years and race. Since years is in fact an ordinal variable, it was scored as the integers 1 to 5 for the years 1988 to 1992. This saves degrees of freedom and helps detect monotonic trends.

The model with a cubic term in years gives an adequate fit to the data (p-value = .112). This model fits the data somewhat better than the model that fits the year as a categorical variable.

The fitted values are the estimates of the conditional probabilities that an officer receives a waiver given year and race. These are plotted by race in Figure 3. From this plot it can be seen that except for 1989 there has been a general decline in the proportion of waivers awarded for each race.

In conclusion, we have accounted for the nature of the paradox stated in the introduction by the use of loglinear analysis after collapsing the data over time. The odds ratio analysis served to support the independence vs. waiver hypothesis at a micro-level, and deeper loglinear modeling can be used to quantify the changes in probabilities as functions of race and time. Based on the data and these models, success in OCS has, in general, declined over time for all racial groups independently of waiver status. There does appear to be a marked difference in the probabilities of success among the racial groups. The final analysis collapses the data over OCS success or failure and treats the use of the waiver. It appears to be diminishing in time but there are some rather prominent separations by race. Some additional study in these areas can be found in [3].

For these data, success in OCS (for those qualifying based upon ASVAB test) has declined over time, and is basically independent of waiver status when conditioned on recorded race. Two hypotheses emerge: the ASVAB contains a racial bias that accepts candidates by race group, leading to uneven success rates; or discrimination occurs in OCS, leading to differential success rates despite equal qualification.

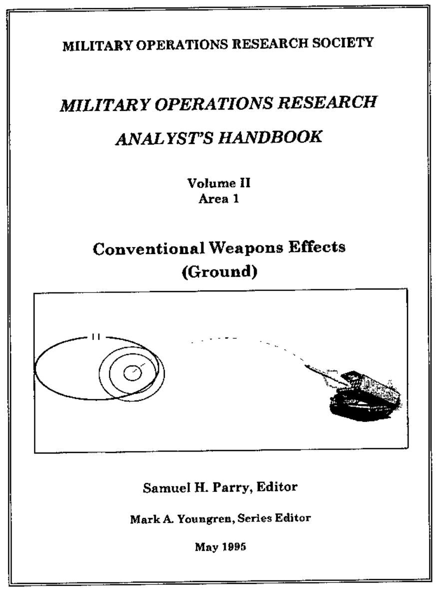
REFERENCES

[1] Agresti, A. (1990). *Categorical Data Analysis*, Wiley.

[2] Davis, L. J. (1986). "Exact Tests for 2 x 2 Contingency Tables," *The American Statistician*, Vol. 40, No. 2.

[3] Harrington, D. F. (1993). "An Equal Opportunity Misconception and the Accession/Selection Paradox," Marine Corps Gazette.

ANALYST'S HANDBOOK SUPPLEMENT NOW AVAILABLE



Area I of Volume II, *Conventional Weapons Effects*, is now available through the MORS office. It is intended to be part of the initial notebook in the *Handbook* series, along with Volume I (*Terrain, Unit Movement, and Environment*) and Volume II, Area II (*Search, Detection and Tracking*) (To be published.)

The *Conventional Weapons Effects* (*Ground*) area was edited by Sam Parry, who has been educating and advising Army OR analysts for many years. It is designed to provide a quick reference for models that represent the effects of a conventional attack against ground targets. It emphasizes the ground battle; publication of algorithms for naval and air engagements is planned for the future. It cannot, of course, include every algorithm used by military OR practitioners, but it is an attempt to characterize those that are most commonly used.

The cost of Volume II Area I is $8.00. As always, Volume I, which includes a three-ring binder for future supplements, is available for $25.00. **Combine both and save: $30.00.** All prices include freight.

Please call the MORS Office for international shipping rates.
VA Residents add 4.5% sales tax.

Volume II Area I: ________ Volume I: ________

Volume I and Volume II, Area I: ________

Total Enclosed: ________ **Make checks payable to MORS**

Name: ______________________________

Shipping Address: ______________________________

____

____

City: ____________ State: ____ Zip Code: ________

If paying by credit card

- ❑ Visa
- ❑ Master Card
- ❑ American Express

Credit Card Number

Expiration Date

Signature

Military Operations Research Society • 101 S. Whiting Street • Suite 202 • Alexandria VA 22304 • (703) 751-7290

Non-Monotonicity, Chaos and Combat Models

J.A. Dewar
J. J. Gillogly
M. L. Juncosa
RAND

INTRODUCTION

Computerized combat models are common tools in the analysis of military strategy, tactics, policy and training [Hughes (1989)]. They are widely used to explore alternative force structures and force employment schemes and are a vital part of the process of choosing which weapon systems to develop and purchase. They belong to a general set of *ad hoc* models whose underlying dynamics are typically nonlinear and have not been subjected to rigorous analysis and testing under controlled conditions. To be sure, some portions of combat (e.g., projectile trajectories, mean-times-between-failures, etc.) may be modeled accurately and tested rigorously under controlled conditions. Other parts, however, (e.g., "leadership," the "fog of war," etc.) are beyond one's ability to represent and test analytically.

The typical model simulates combat between opposing forces at some level of abstraction. No combat model is seriously expected to be precisely predictive of actual combat outcomes. It is common, however, to expect models to be relatively predictive. That is, if a capability is added to one side and the battle is refought, the difference in battle outcomes is expected to reflect the contribution of the added capability.

Models used for comparative purposes, then, carry the implicit assumption that combat is what we will call (borrowing the mathematical term) *monotonic*, in that adding more capabilities (only) to one side will lead to at least as favorable a combat outcome for that side. The model's outcomes are then interpreted in light of that implicit assumption and are generally disbelieved if they exhibit non-monotonicities. Although non-monotonic behavior is not uncommon in combat models, modelers generally treat it as anomalous and either "fix" the model until the non-monotonicity disappears or ignore the non-monotonic outcomes. By their nature, however, combat models tend to be large and complex and can take several hours of computer time to produce the results of a single simulated battle. Because of this, the "fixes" tend to be for a specific observed non-monotonicity, and little is done to explore the model in general for non-monotonicities.

On the other hand, non-monotonicities are known to be caused by a wide variety of mechanisms. Many combat models, for example, contain stochastic variables, and random variations from battle to battle can clearly cause some non-monotonicities to creep in. Even in deterministic models with no random components or one whose random variables are replaced by their mean values non-monotonicities can arise from several well known causes. If properly dealt with, these causes and their resultant non-monotonicities can typically be eliminated.

We were interested in a different source of non-monotonicities—specifically non-monotonicities related to modeled command decisions on reinforcements. Such decisions, based on the state of the battle, introduce mathematical nonlinearities into a model. These nonlinearities can be shown to be the cause of non-monotonicities in the model's outcomes. Our interest, however, lay in the fact that the same nonlinearities can also lead to mathematical chaos. We wanted to know if, in fact, chaos was present, and what (if any) relation it had to the observed non-monotonicities. But what is chaos and why might it be important to combat models?

CHAOS

The advent of video-display microcomputers has greatly increased the visibility and understanding of a class of physical and mathematical processes identified with chaotic behavior or chaos (see Gleick 1987 or Stewart 1989). Chaos has now been recognized and investigated in a wide variety of disciplines including weather forecasting [Lorenz 1963, Palmer 1989, Pool 1989e], chemical reaction kinetics [Rehmus 1985, Scott 1989], population dynamics [May 1976, 1989], planetary orbits [Murray 1989], the arms race [Saperstein 1984, Grossmann 1989], epidemiology [Pool 1989], the oscillations of atomic particles [Hoffnagle 1988], economic prices [Jensen 1984, Nash 1988] and neural networks [Derrida 1988].

While no simple, universally accepted definition of chaotic behavior exists, chaos is characterized by unpredictable, random-looking behavior over long periods and extreme sensitivity to current or initial conditions. Chaotic behavior is not necessarily a product of random impulses but can be implicit in the deterministic equations modeling the process with no stochastic elements in them and can be observed in the behavior of the process. The significance of chaos in

Application Areas:
Verification, validation; conventional force analysis

OR Methodologies:
Nonlinear dynamical systems

mathematical simulations (and in the physical process being simulated) is that the outcomes do not settle out to some steady state or even a predictable cycle; they must be fully calculated through all of their iterations; and they are so sensitive to initial conditions as to make each simulation a unique path with little or no relationship to its neighbor only slightly removed.

Combat simulations, particularly those for ground combat, are candidates for chaotic behavior because they often involve nonlinear equations that are iterated many times over the course of a battle and, with reinforcements, combat simulations contain both forcing and damping behavior common to chaotic systems. Further, combat models often exhibit non-monotonicities indicative of a sensitivity to small changes. These facts suggest that non-monotonicities occasionally seen in combat models might be related to chaos. If they are, they may be more widespread than they are generally given credit for, and they may represent *inherent* rather than anomalous behavior in such models. This leads to the possibility that some aspects of the simulation (and of the battle being simulated) are simply not predictable or are extremely sensitive to the initial conditions or intervening events. One must then question the validity of comparisons made with such a model.

Some work has been done on relating chaos and combat models. Some analysts, for example, believe they have observed chaotic behavior in large combat simulations, specifically in VIC (Sandmeyer 1988). It is difficult to prove that the observed phenomena are indeed evidence of chaos rather than simple sensitivity, noise from rounding errors, or some other cause because these simulations are so complex, they are neither transparent nor, because of the time required to run them, easily mapped over a wide range of conditions. Work at Oak Ridge National Laboratories is ongoing in trying to model combat through partial differential equations [Protopopescu 1989] and, in the course of that work, they have studied chaos in combat models of that type. In a related field, Saperstein, Grossmann and Mayer-Kress have written on chaos and the arms race [Saperstein 1984, Grossmann 1989].

APPROACH

In general, there are several challenges in investigating non-monotonicities and chaos in combat models. First, one needs to be assured that the non-monotonicities under investigation are not the result of causes other than the nonlinearities associated with the potential chaos. Second, there are several definitions of chaos to be found in the literature (see Collet 1980: p. 15 for examples) so one must be careful in choosing a definition commensurate with combat modeling. Third, chaos is a long term behavior of dynamical systems and typical combat models are run for a relatively small number of time steps. Behavior that *appears* chaotic in a given combat model run requires careful analysis of the underlying equations to demonstrate that they satisfy the requirements for mathematical chaos. Finally, if chaos is present in the underlying dynamical system, one needs to make clear the relationship between that chaos and misbehavior in (finite) realizations of the system and their stopping conditions.

These investigations required a combat model that was both amenable to mathematical analysis and that would permit literally millions of runs during the course of the work. If a connection could be made between chaos and non-monotonicities in *any* combat model, this would serve as an "existence theorem" for such behavior. Making a connection between the behavior of that model and other combat models would be a subsequent step. Taking the first step required a very simple model, but one with at least some semblance of reasonability.

The basic model we created to serve our purposes is shown in Table 1. B_n and R_n represent troop strengths of Blue and Red at time n. For each battle, each side starts with a fixed number of troops, B_0 and R_0. All these troops are presumed to be in contact and fighting continuously. The dynamics of the battle are described by the attrition equations in Table 1 modified by the incremental reinforcements whenever the reinforcement thresholds are crossed. The attrition coefficients were chosen as powers of 2 in order to aid in computational precision. The time step of the model is inherent in the selection of the attrition coefficients (and reinforcement delays) and in this case it has been chosen to

	Blue	Red
Initial Troop Strength	Variable	Variable
Combat Attrition Calculations	$B_{n+1} = B_n - \frac{R_n}{2048}$	$R_{n+1} = R_n - \frac{B_n}{512}$
Reinforcement Thresholds	$\frac{R_n}{B_n} \geq 4$ <u>or</u> $B_n < .8\ B_0$	$\frac{R_n}{B_n} \leq 2.5$ <u>or</u> $R_n < .8\ R_0$
Reinforcement Block Size	300	300
Maximum Allowable Reinforcement Blocks	5	5
Reinforcement Delay (time steps)	70	70
Withdrawal Thresholds	$\frac{R_n}{B_n} \geq 10$ <u>or</u> $B_n < .7\ B_0$	$\frac{R_n}{B_n} \leq 1.5$ <u>or</u> $R_n < .7\ R_0$

TABLE 1
Simple Combat Model

represent about half an hour of simulated battle per step.

In the example of Table 1, the Blue "commander" calls for reinforcements whenever the Red-to-Blue force ratio exceeds four *or* whenever his force drops below 80% of his initial troop strength. In this model, after he has called for reinforcements, he may not call for more until those he just called for arrive. All reinforcements are delayed by the number of time steps specified by "Reinforcement Delay." The 70 time step delay in Table 1 represents about 35 hours in the simulated battle. The reinforcements come in blocks and the commander has a maximum number he can call for.

Not all combat models produce a loser or winner, but all have stopping criteria. In this model, the stopping criteria always cause one or both sides to withdraw, thus ending the battle. In Table 1, the Blue commander will withdraw (thus "losing") if the Red-to-Blue force ratio exceeds 10 or if his force is below 70% of his initial troop strength. In what follows, Red will be declared the "winner" of this battle unless he, too, withdraws on the same time step. In that case, the battle is a "draw."

Two examples will serve both to illustrate the behavior of this simple model and to dramatize the meaning of non-monotonicities. Figure 1 represents 2001 battles using the parameters in

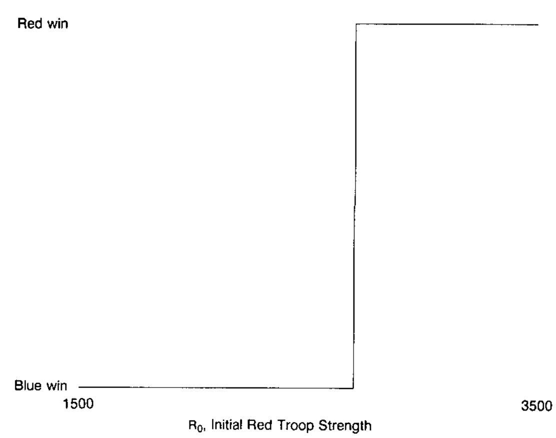


FIGURE 1
Monotonic Behavior

Table 1, with Blue's initial troop strength fixed at 839 troops and Red's ranging from 1500 to 3500. The outcomes represent monotonic behavior in that once Red wins (as he does when starting with 2696 troops), adding more Red troops at the start of the battle doesn't change the outcome. If, however, Blue's initial troop strength is fixed at 500 troops and another series of battles is run, the outcomes are as shown in Figure 2. Here the battles range over Red initial troop strengths from 700 to 1800. Outcomes in this region exhibit seriously non-monotonic behavior in that Red can win when starting with as few as 884 troops,

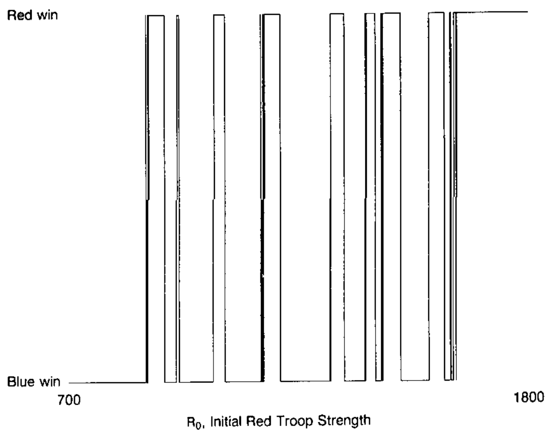


FIGURE 2
Non-Monotonic Behavior

can lose when starting with as many as 1623 troops, and suffers a surprising number of reversals of fortune in between. In this region, the outcome of an individual battle is not very predictive of the outcome of a "nearby" battle (i.e., one starting with nearby initial troop strengths). Said another way, the outcome of a battle in that region, from 700 to 1800 starting troops, is sensitive to small variations in Red's initial strength.

Figures 1 and 2 are one-dimensional in that they illustrate the outcomes of battles as one varies the number of Red troops. More generally we will be interested in two-dimensional pictures in which the outcomes are plotted along both Blue and Red starting troop dimensions. Figure 3 is an example of such a plot, with black points representing Red wins and white points representing Blue wins. The points plotted are the outcomes of battles with starting troops that are multiples of 5. Some detail is lost in so doing, but by this compromise, we get to see more of the outcome space without seriously affecting the indications of non-monotonicity. Figures 1 and 2 represent horizontal slices in this figure, as shown. This figure emphasizes the surprising extent over which non-monotonicities may be found.

As simple as the model in Table 1 is, there are 18 different parameters that may be varied making the input space of this model 18 dimensional. Understanding what happens to this model requires understanding what happens across all 18 dimensions. Because we were inter-

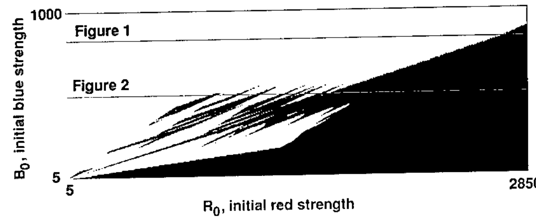


FIGURE 3
Simple Model Outcomes in Two Dimensions

ested in understanding the non-monotonic behavior and because exploring 18 dimensions is a formidable computational task, we also looked at two subsets of this model. If one deletes the four thresholds dealing with force ratios in Table 1, one has a smaller model that we refer to as the "attrition-only" model. Deleting, instead, the four attrition thresholds yields a "force-ratio-only" model. These each have input spaces of "only" 14 dimensions. Not only does this reduce (somewhat) the computational formidability of the task, but it allowed us to ask and (partially) answer some questions about what happens to non-monotonicity when one combines thresholds in a simple combat model.

Having chosen the model, the challenges were to isolate the source of non-monotonicities to the nonlinearities in the command decision to reinforce, *prove* the underlying equations satisfy the definition of mathematical chaos, connect the non-monotonicities with the underlying chaos, and generalize the results to larger models to the extent possible.

RESULTS

Eliminating Extraneous Causes of Non-monotonicity

There are a variety of modeling actions that can lead to non-monotonicities in the model's outcomes. Among well known causes of non-monotonicities are time step granularity problems, delayed feedback effects, a variety of roundoff/precision problems, the effects of random variables and those of smoothing or time-averaging. Our model is deterministic and contains no smoothing or time-averaging, so the last two of these can be eliminated as potential

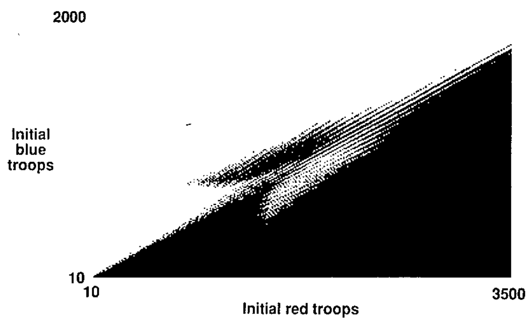


FIGURE 4
Inappropriate Time Step Size

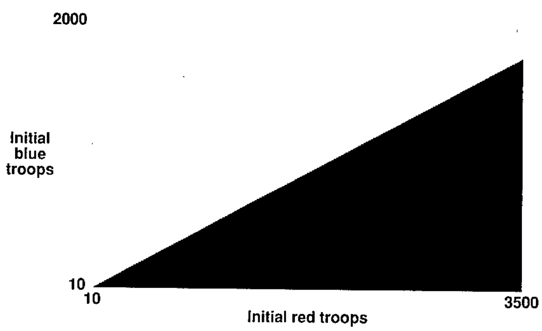


FIGURE 5
Appropriate Time Step Size

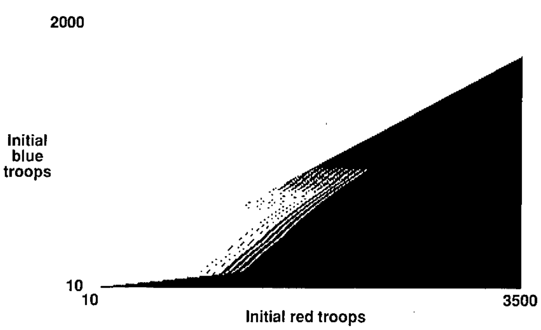


FIGURE 6
No Reinforcement Delays

problems. In our simple model, time step granularity is a particular concern because the model is basically a discrete approximation to an underlying differential equation. If the time step is taken too large, non-monotonicities can occur. Figure 4 is an example of our model (in an otherwise monotonic region) with the time step (inherent in the attrition and delay coefficients) taken eight times as large as in Table 1. Figure 5 shows the same region with the time step as in Table 1 (note that the choice of the time step is independent of the situation being modeled and depends only on the mathematical behavior of the model). The general size of the attrition coefficients was chosen to minimize time step granularity problems without unduly slowing the model's turnaround time. Further, the coefficients were chosen to be powers of 2 in order to minimize precision / roundoff problems.

The remaining potential extraneous cause of non-monotonicities is the reinforcement delay. Delaying the reinforcements after the decision to call them is reasonable from a combat modeling standpoint, but sets up the possibility of a feedback loop that engineers are well aware can cause instabilities in the system. The solution to this problem illustrates the general problem of settling mathematical difficulties in a combat model. Setting the reinforcement delay to zero will eliminate this problem, but makes the model a bit unrealistic from a practical standpoint—reinforcements don't get there instantaneously. This problem can be finessed to some extent by suggesting that the decision was made based on projections of the state of the battle so that the reinforcements were scheduled to arrive when the threshold for calling them was breached. This is an arguable modeling rationalization, but for our purposes, it suffices simply to look at the model with zero delay. That it still produces non-monotonicities is exhibited by Figure 6.

Having controlled for the better-known causes of non-monotonicities, the behavior of the resulting model should be almost entirely due to the form of the model and its dynamics. A more positive indication that the remaining non-monotonicities are due to the nonlinearities introduced by the reinforcement decision can be seen by removing the reinforcement decision and seeing what happens. This can be done by introducing the reinforcements as a function of time only. This is often called "scripting" the reinforcements and is a common modeling trick for getting rid of non-monotonicities. In our case, scripting the reinforcements in Figure 6 leads to Figure 7. In fact, it can be shown that scripting the reinforcements here produces a finite linear model that cannot have non-monotonicities—which explains why scripting the reinforcements gets rid of our problem (but at the cost of verisimilitude!).

	Blue	Red
Initial troop strength	40	80
Combat attrition calculation	$B_{n+1} = B_n - 0.005\,R_n$	$R_{n+1} = R_n - 0.020\,B_n$
Reinforcement thresholds	$\frac{R_n}{B_n} \geq 4$	$\frac{R_n}{B_n} \leq 2.5$
Reinforcement block size	10	10
Maximum allowable reinforcement blocks	Unlimited	Unlimited
Reinforcement delay (time steps)	0	0
Withdrawal thresholds	None	None

TABLE 2
Modified Model for Investigating Chaos

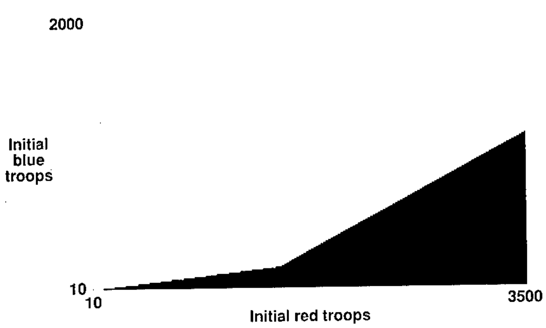


FIGURE 7
Scripted Reinforcements

Proving There is Chaos in the Underlying Model

From a dynamics standpoint, the model is "artificially" halted either by the restrictions on the number of available reinforcements or by the stopping conditions. If both of these restrictions are removed, the asymptotic behavior of the underlying dynamic process can be studied. To make this point clear, Table 2 shows the model that was actually tested for chaos. Note this is the "force-ratio-only" model and that the attrition coefficients are slightly different (the latter is unimportant, because the system was studied analytically, not computationally). The long term behavior of this model is pictured in Figure 8 (which was obtained mathematically as well as empirically). In the "phase space" of remaining Red and Blue troops, the battle loops forever in the "attractor" region pictured. The question of chaos rests on whether or not that attractor is chaotic.

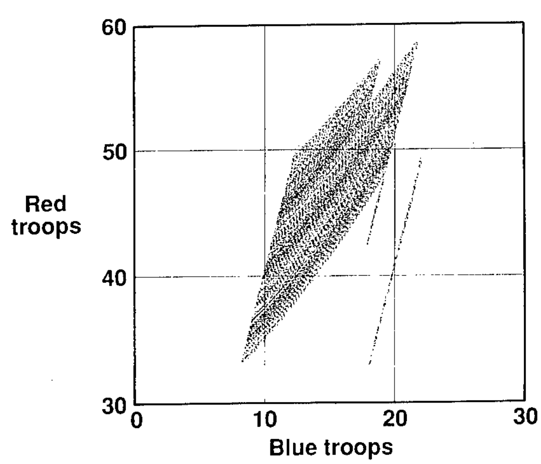


FIGURE 8
Phase Space Attractor

There is still no generally accepted definition of chaos (see Liechtenberg 1983: p. 213 for a list of six characteristics that various authors have used). Even for continuous flows and maps, there are several definitions in the literature. Our case is complicated in that, even though the equations for our model are linear where they are continuous, they are only piecewise continuous with the discontinuities at the reinforcements. There are discontinuous maps which satisfy published definitions of chaos (the baker transformation mentioned in Devaney (1989) is a classical example), but ours does not in the strict sense. To understand our discussion of chaos, we begin with the definition for chaos found in Devaney (1989).

Let V be a set. *If*: V -> V is said to be chaotic on V if

1. *If* has sensitive dependence on initial conditions.
2. *If* is topologically transitive.
3. periodic points of *If* are dense in V.

Devaney states: "To summarize, a chaotic map possesses three ingredients: unpredictability, indecomposability, and an element of regularity. A chaotic system is unpredictable because of the sensitive dependence on initial conditions. It cannot be broken down or decomposed onto two subsystems (two invariant open subsets) which do not interact under *If* because of topological transitivity. And, in the midst of this random behavior, we nevertheless have an element of regularity, namely the periodic points which are dense." Elsewhere, having an invariant density is an important aspect of the definition of chaos. While our modified model doesn't strictly meet these chaos criteria it does have the following properties:

1) Generally sensitive dependence on initial conditions.
2) Topological transitivity.
3) An infinite number of periodic points.
4) An invariant density (with gaps).

While some of these properties are satisfied in a restricted sense, our model clearly satisfies the *spirit* of the definition of chaos in the descriptive sense that Devaney uses above. It is inappropriate at this point to introduce yet another definition of chaos (for piecewise continuous maps), but we consider our piecewise continuous map to be chaotic because it satisfies the four conditions above.

Connecting Chaos with Non-monotonicities

It is necessary to take some care when discussing the relation between chaos in the modified infinite model and non-monotonicities in the simple, finite combat model. In what sense can we say that the chaos in the infinite model is linked to non-monotonicities in the finite model? If one takes the chaos away from the infinite model, and non-monotonicities disappear from the finite model, this is strong evidence they are linked. That this is the case here was demonstrated in Figures 6 and 7.

It is important to note that the models used in Figures 6 and 7 both had stopping criteria based on the state of the battle. It is clear in this case that any nonlinearities they introduce are not causing or remedying the non-monotonicities, but what effect do the stopping criteria have on the finite model in general?

There are stopping criteria in any finite model. Practically speaking, as soon as either side has fewer than, say, one remaining troop, it must stop fighting and the attrition equations generally ensure that eventually this must happen. More commonly, however, battles (both real and simulated) are stopped long before annihilation of one side. In some cases, the battle will be stopped independently of its progress (e.g., after a certain amount of simulated or computer time has elapsed). In other cases, there will be stopping criteria established that are a function of the state of the battle.

Whatever the stopping criteria, though, non-monotonicities are associated most generally with a further evaluative mapping from the final state of the battle to an ordered set of states. A popular such set is the binary set "win" and "lose", but there are a variety of other such sets ranging from territory won or lost to multi-dimensional measures of materiel used, etc. As long as the evaluative mapping is ordered, then a non-monotonicity is any unexpected reversal in outcome associated with a given change in inputs.

In our case, the non-monotonicity is a

change from one side winning a battle to that side losing given that the only change in inputs is an increase in that side's initial troop strength. To see most clearly that it is the nonlinearities associated with the reinforcement decision that are necessary to cause non-monotonicities in our finite battles to appear, we looked at what makes for a finite model in our case. There are two basic differences between the modified model in Table 2 and the finite force-ratio-only submodel of the original model Table 1 (with zero delay): 1) There are stopping criteria in the finite model and 2) only a finite number of reinforcements can be called upon in the finite model. To eliminate the stopping criteria as a potential source of non-monotonicities, we got rid of them and stopped the battle only with the "natural" stopping condition—when one side was annihilated. To ensure finite battles, we restricted the number of reinforcements available to each side. Any non-monotonicities in this model must be related to the fact that the reinforcement schedules are potentially different from battle to battle because of the reinforcement heuristic. There were still non-monotonicities. That is, the (chaotic) nonlinearities introduced by the reinforcement heuristic produced unexpected changes in the fundamental progress of the battle. In that sense it must be said that the chaotic nonlinearities lead to "non-monotonic" behavior in the model. Non-monotonicities in model outcomes are thus a necessary corollary.

Implications for Larger Models

There are two sobering generalizations that can be made about larger, more complex models based on our work. Both of them deal with adding state dependent thresholds to a model. The first deals with the behavior of the model itself and the second with the ambiguity region of a model.

Figure 9 shows the results of our attrition-only model for a given set of parameters. Its behavior is reasonably monotonic. With the same set of parameters, but using force ratios instead of attrition for both reinforcement and withdrawal thresholds, one gets Figure 10. It, too, is reasonably monotonic. But these are just the attrition-only and force-ratio-only submodels of the model in Table 1. That is, if we added force

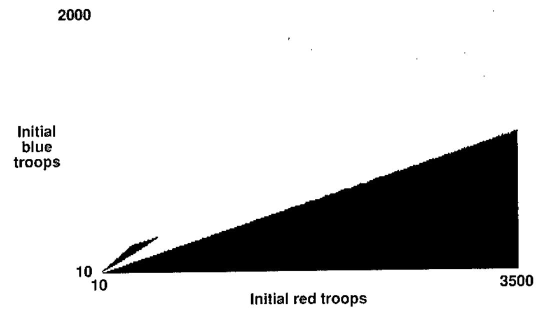


FIGURE 9
Simple Model: Attrition Thresholds Only

ratio thresholds to the attrition-only model of Figure 9 or attrition thresholds to the force-ratio-only model of Figure 10, we get Figure 3.

In other words, here is an example where either submodel is reasonably well behaved. If, however, we add the two submodels together the result is seriously non-monotonic behavior. This is a counterexample to a suggestion we have heard from several modelers that perhaps having many thresholds would "wash out" the undesired effects of any given threshold. Adding a threshold *can*, perhaps, improve the behavior of a model (possibly even by *moving* the undesirable behavior out of a given region), but it is clear from this example that adding a threshold can also, demonstrably, worsen the behavior of the model.

Another hypothesis about adding thresholds is that, perhaps, adding a threshold might worsen a model's behavior for a given set of parameter values, but that, overall, it would be shrinking the area in which non-monotonicities might occur. In other words, perhaps it is worsening the situation in a local region, but adding the threshold shrinks the region in which non-

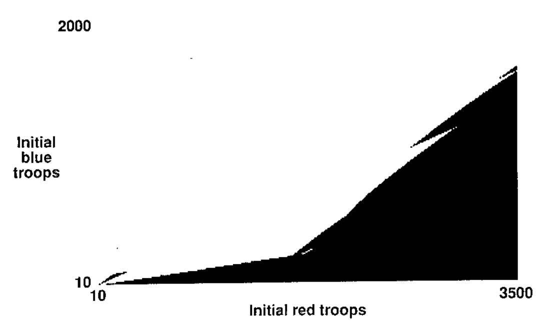


FIGURE 10
Simple Model: Force Ratio Thresholds Only

monotonicities might occur, and that adding enough thresholds would shrink that region to a manageable or insignificant size.

It is possible to describe, analytically, the regions outside of which non-monotonicities *cannot* occur for our simple models. For each of the force-ratio-only and attrition-only models, there is an optimal reinforcement strategy for both Red and Blue. Further, because it is the same strategy for both submodels, there is an optimal reinforcement strategy for the combined model.

It can be shown that the optimal reinforcement strategy for both the force-ratio-only and attrition-only models and for both Red and Blue is to have all their reinforcements immediately at the beginning of the battle. If one side (say Blue) has all its reinforcements immediately and the other side never calls for reinforcements then the outcome for Blue for any starting conditions will be as good as Blue can do. Note that because there are no reinforcement calls, there is no chance for non-monotonicity and the resulting picture over a range of starting conditions will have a clean demarcation line between Blue and Red wins. That line represents the edge of the area outside of which non-monotonicities cannot occur. Specifically, the area of Red wins represents an area of assured Red wins no matter what the reinforcement strategies employed by the two commanders. By reversing the situation (Red has all its reinforcements immediately and Blue never has any) another line can be generated outside of which there can be no non-monotonicities in the Blue direction. Combining the lines on the same picture defines the limits of the area within which non-monotonicities *can* occur. We call this area the *ambiguity region,* or the region within which there *can* be non-monotonicities.

Figure 11 shows the ambiguity regions drawn onto a plot that contains non-monotonicities. Figure 12 shows two separate ambiguity regions for attrition-only and force-ratio-only submodels that otherwise have the same model parameters. The grey line shows the ambiguity region of the model that combines the two submodels. The significance of this figure is not the exact size of the ambiguity region of the combined model, but rather the fact that it is *not* the union or the intersection of the ambiguity regions of the submodels. Many of the examples

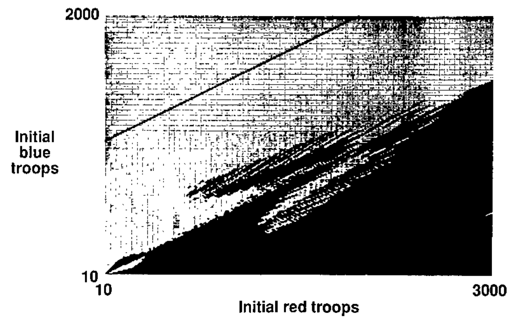


FIGURE 11
Ambiguity Region

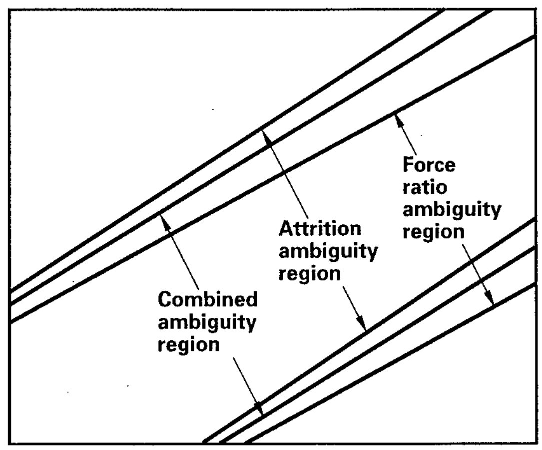


FIGURE 12
Combining Ambiguity Regions

we ran in this part of the research had the union of the two constituent regions as the ambiguity region of the combined model. However, since there are times when it is neither the union nor the intersection, the general behavior of ambiguity regions in large, complex models is likely to be complex and difficult to gauge.

DISCUSSION

For an important class of combat phenomena—reinforcement decisions based on the state of the battle—we have shown that modeling this behavior can introduce nonlinearities that can lead to chaotic behavior in the dynamics of computerized combat models. That is, in a simple combat model without stopping conditions, we have shown for a specific decision—when to call in battle reinforcements—based on the state of the battle—specifically, on the ratio of the

opposing forces strengths—that the underlying dynamics of the model satisfy four mathematical conditions characteristic of chaotic systems. Further, we have shown that, for a variety of stopping conditions, the chaotic dynamics of the underlying system give rise to non-monotonicities in model outcomes.

Because of the chaotic underlying dynamics, the sensitivity to initial conditions associated with the nonlinear reinforcement heuristics will appear, for example, even if the dynamical system is solved exactly. The "misbehavior" of this model is structural rather than computational, it is in the nature of the phenomenon being modeled—decisions based on the state of the battle. We have shown, further, that this structural misbehavior can lead to non-monotonicities in the outcomes of the model, and that the non-monotonic behavior can be spread over wide areas of the input parameter space. In this sense, then, decisions based on the state of the battle can be seen to "cause" widespread non-monotonicities in the outcomes of the model.

To the extent that monotonic behavior is important for the uses of a given model, non-monotonic behavior is bad. This almost tautological statement appears to be underappreciated in the modeling community. Consider a model that is to be used for comparative purposes; that is, a model to be used to compare two or more systems, force structures, doctrines, or other alternatives. Valid comparisons require that the model outcomes accurately reflect at least the relative contributions to the battle of the competing alternatives. Suppose, instead, the model outcomes are reflecting a combination of the contributions of the competing alternatives *and* non-monotonicities from the underlying dynamics of the combat model. In order for the comparisons to be valid in this case, it must be that the observed non-monotonicities reflect what would happen in real battle. This is a crucial point: while it is conceivable that they might, unless it has been validated that they *do*, the comparisons are questionable. Since no combat model has ever been (or *could ever be*) validated in this sense, comparisons among competing alternatives are invalid to the extent that the model being used exhibits non-monotonic behavior in the region of the comparisons.

Models used for comparative purposes, then, must be monotonic in order for the comparisons to be arguably valid. How might the non-monotonicities observed in our model be eliminated? We have shown that if the reinforcement heuristic is not a function of the state of the battle (e.g., if it is "scripted" as a function of time), the nonlinearities, the chaos and the non-monotonicities all disappear. This also removes the verisimilitude of having the decision made based on the progress of the battle—as it is done in real life. For those comparisons, however, that are not critically dependent on such verisimilitude, scripting the reinforcements in the model will eliminate non-monotonicities associated with nonlinear decision heuristics.

The only other mitigation supported by our research is that of exhaustively verifying that the model does not exhibit non-monotonic behavior in the subspace of input parameter space of interest. That is, we have seen that, while non-monotonicities can be quite extensive in some regions of input parameter space, other regions are quite monotonic. If one can verify that the model is, indeed, monotonic over the region of input parameter space in which it will be exercised, then, comparisons in that region are arguably valid with respect to the model's underlying dynamics.

Put in other words, we have demonstrated that a combat model with a single decision based on the state of the battle, no matter how precisely computed, can produce non-monotonic behavior in the outcomes of the model and chaotic behavior in its underlying dynamics. Working models, however, have not a single such decision, but a number of such decisions ranging from dozens to thousands. What can be said of their behavior in this regard? We have shown that adding another decision based on some state of the model can worsen any observed non-monotonicities, and that the area in which these non-monotonicities can occur does not necessarily shrink when the decision is added.

In conclusion, then, when comparisons of strategy, tactics or systems are based on a combat model that depends on monotonic behavior in its outcomes, modeling combat decisions based on the state of the battle must be done very carefully. Such modeled decisions can lead to non-monotonic and chaotic behavior and the only sure ways (to date) of dealing with that behavior are either to remove the modeled decisions or to

verify that the model is monotonic in the region of interest. Today the focus of combat modeling is shifting from Europe to other theaters and new models of combat will have to be developed. Matters of non-monotonicity and chaos should be addressed early in the design phases of these new models.

REFERENCES

Azmy, Y. Y. and V. Protopopescu, *Two Dimensional Maps Generated by Competitive Systems*, ORNL/TM-11026, Oak Ridge National Laboratory, 1989.

Barnsley, Michael, *Fractals Everywhere*, Academic Press, California, (1988).

Benzinger, Harold E., Scott A. Burns and Julian I. Palmore, "Chaotic Complex Dynamics and Newton's Method," *Physics Letters A*, **119** (9), pp. 441-446, 1987.

Burns, Scott A. and Julian I. Palmore, "The Newton Transform: An Operational Method for Constructing Integrals of Dynamical Systems," *Physica D*, **37**, pp. 83-90, 1989.

Campbell, David K. and Gottfried Mayer-Kress, *Nonlinearity and Chaos: The Challenges and Limitations of Mathematical Modeling of Environmental and Socio-Political Issues*, Center for Nonlinear Studies, Los Alamos National Laboratories, 1989.

Cipra, Barry A., "Computer-Drawn Pictures Stalk the Wild Trajectory," *Science*, **241**, pp. 1162-1163, 1988.

Cipra, Barry A., "Image Capture by Computer," *Science*, **243**, pp. 1288-1289, 1989.

Collet, Pierre and Jean-Pierre Eckmann, *Iterated Maps on the Interval as Dynamical Systems*, Birkhauser, Basel (1980).Crutchfield, James P., "Quantifying Spatial Chaos," *Research and Development*, pp. 13-17, 21 Nov 1989.

Culioli, J.-C. and V. Protopopescu, *Bifurcating Optimization Algorithms and their Possible Application*, ORNL/TM-10976, Oak Ridge National Laboratory, 1988.

Cvitanovic, Predrag, *Universality in Chaos*, Adam Hilger, Bristol (1984).

Derrida, B. and R. Meir, "Chaotic Behavior of a Layered Neural Network," *Physical Review A*, **38** (6), 1988.

Devaney, Robert L., *An Introduction to Chaotic Dyamical Systems*, Addison-Wesley, New York (1989).

Dockery, J. T. and A. E. R. Woodcock, "Models of Combat with Embedded C^2II: Catastrophe Theory and Chaotic Behavior," *International CIS Journal*, **2** (4), pp. 17-51, 1988.

Eckmann, J. P. and D. Ruelle, "Ergodic Theory of Chaos and Strange Attractors," *Reviewers of Modern Physics*, **57** (3), Part 1, pp. 617-656, 1985.

Froehling, Harold, J. P. Crutchfield, Doyne Farmer, N. H. Packard and Rob Shaw, "On Determining the Dimension of Chaotic Flows," *Physica 3D*, pp. 605-617, 1981.

Gallavotti, Giovanni and Donald S. Ornstein, "Billiards and Bernoulli Schemes," *Communications in Mathematical Physics*, **38**, pp. 83-101, 1974.

Geweke, John, *Inference and Forecasting for Chaotic Nonlinear Time Series*, (Duke University, North Carolina, 1989).

Gleick, James, *CHAOS, Making a New Science*, Viking, New York (1987).

Gordon, Theodore J. and David Greenspan, "Chaos and Fractals: New Tools for Technological and Social Forecasting," *Technological Forecasting and Social Change*, **34**, 1988.

Grossman, Siegfried and Gottfried Mayer-Kress, "The Role of Chaotic Dynamics in Arms Race Models", *Nature*, **337**, pp. 701-704, 1989.

Guckenheimer, John and Philip Holmes, *Nonlinear Oscillations, Dynamical Systems, and Bifucations of Vector Fields*, Springer-Verlag, New York (1986).

Hoffnagle, J., et. al., "Order-Chaos Transition of Two Trapped Ions" *Physical Review Letters*, **61**, 255, 1988.

Homer-Dixon, Thomas F., "A Common Misapplication of the Lanchester Square Law," *International Security*, **12** (1), pp. 135-139, 1987.

Hughes, Wayne P., Jr., *Military Modeling*, Military Operations Research Society (1989).

Jensen, Roderick V., "Classical Chaos," *American Scientist*, **75**, pp. 168-181, 1987.

Jensen, R. V. and R. Urban, "Chaotic Price Behavior in a Nonlinear Cobweb Model, *Econ. Lett.*, **15**, 1984.

Kan, Ittai, "Strange Attractors of Uniform Flows," *Transactions of the American Mathematical Society*, **293** (1), pp. 135-159, 1986.

Kloeden, P. E., "Chaotic Difference Equations in R", *Journal Australian Mathematical Society*, **31**, pp. 217-225, 1981.

Lepingwell, John W. R., "The Laws of Combat?," *International Security*, pp. 89-133.

Li, T.-Y., and J. A. Yorke, "Period Three Implies Chaos," *Amer. Math. Monthly*, **82**, pp. 985-992, 1975.

Lichtenberg, A. J., and M. A. Lieberman, *Regular and Stochastic Motion*, Springer-Verlag, New York (1982).

Lorenz, E., "Deterministic Nonperiodic Flow," *J. Athmos. Sci.*, **20**, pp. 130-141, 1963.

Marotto, Frederick R., "Snap-Back Repellers Imply Chaos in R," *Journal of Mathematical Analysis and Applications*, **63**, pp. 199-223, 1978.

May, R. M., "Simple Mathematical Models with Very Complicated Dynamics," *Nature*, **261**, 1976.

May, Robert, "The Chaotic Rhythms of Life," *New Scientist*, pp. 37-41, 18 Nov 1989.

Mayer-Kress, Gottfried, *A Nonlinear Dynamical Systems Approach to International Security*, Proceedings of AAAS Conference, San Francisco, Jan 1989.

McCauley, Joseph L., Jr. and Julian I. Palmore, "Computable Chaotic Orbits," *Physics Letters A*, **115**, (9), 1986.

Moore, Cristopher, "Unpredictability and Undecidability in Dynamical Systems," *Physical Review Letters*, **64** (20), pp. 2354-2357, 1990.

Mullin, Tom, "Turbulent Times for Fluids," *New Scientist*, pp. 52-55, 11 Nov 1989.

Murray, Carl, "Is the Solar System Stable?," *New Scientist*, pp. 60-63, 25 Nov 1989.

Nash, David H. and Brian P. McBride, *Fractals, Chaos Theory and Stock Averages*, Eighth International Symposium on Forecasting, Amsterdam, June 1988.

Ornstein, D. S., "Ergodic Theory, Randomness, and 'Chaos'," *Science*, **243**, pp. 182-187, 1989.

Palmer, Tim, "A Weather Eye on Unpredictability," *New Scientist*, pp. 56-59, 11 Nov 1989.

Palmore, Julian J. and C. Herring, "Computer Arithmetic, Chaos and Fractals," *Physica D*, **42**, pp. 99-110, 1990.

Palmore, Julian I. and Joseph L. McCauley, "Shadowing by Computable Chaotic Orbits," *Physics Letters A*, **122** (8), pp. 399-402, 1987.

Palmore, Julian I., Scott A. Burns, and Harold E. Benzinger, "Ecology Models and Newton Vector Fields," *Mathematical Biosciences*, **90**, pp. 221-232, 1988.

Percival, Ian, "Chaos: A Science for the Real World," *New Scientist*, pp. 42-47, 21 Oct 1989.

Perkers, Kelly, Dan Burnett and Stavros Busenberg, *Theory of Chaos Applied to Combat Models*, Mathematics Clinic Report, Harvey Mudd College, May 1990.

Pool, Robert, "Chaos Theory: How Big an Adventure?," *Science*, **245**, pp. 26-28, 1989.

Pool, Robert, "Ecologists Flirt with Chaos," *Science*, **243**, pp. 310-313, 1989.

Pool, Robert, "Is It Chaos, or Is It Just Noise?," *Science*, **243**, pp. 25-28, 1989.

Pool, Robert, "Is It Healthy to Be Chaotic?," *Science*, **243**, pp. 604-607, 1989.

Pool, Robert, "Is Something Strange About the Weather?," *Science*, **243**, pp. 1290-1293, 1989.

Pool, Robert, "Seeing Chaos in a Simple System," *Research News*, **241**, pp. 787-788, 1988.

Protopopescu, V. and R. T. Santoro, J. Dockery, "Combat Modeling with Partial Differential Equations," *European Journal of Operational Research*, **38**, pp. 178-183, 1989.

Protopopescu, V., R. T. Santoro, and Y. Y. Azmy, *PDE Models for Combat: Potential Applications to LIC*, Oak Ridge National Laboratory, (forthcoming).

Protopopescu, V., R. T. Santoro, R. Cox and P. Rusu, *Combat Modeling with Partial Differential Equations: The Bidimensional Case*, ORNL/TM-11343, Oak Ridge National Laboratory, 1990.

Rehmus, Paul and John Ross, "Periodically Perturbed Chemical Systems," *Oscillations and Traveling Waves in Chemical Systems*, pp. 287-332, John Wiley, New York (1985).

Rydell, Peter and Richard Stanton, *Frequency and Examples of Nonconvergence in Moscow*, 33 pp. (1989).

Rydell, Peter, *Cycles, Chaos and Cliffs in the Moscow Model*, 23 pp. (1989).

Sandmeyer, R. S., "SIMTECH Project: Application of Supercomputers to Division/Corps Level Combat Simulation," Presentation at VIC Model Users Group meeting, White Sands Missile Range, NM, Oct 1988.

Saperstein, Alvin M. and Gottfried Mayer-Kress, "Chaos Versus Predictability in Formulating National Strategic Security Policy," *Am. J. Phys.*, **57** (3), pp. 217-223, 1989.

Saperstein, Alvin M. and Gottfried Mayer-Kress, "A Nonlinear Dynamic Model of the Impact of SDI on the Arms Race", *J. Conflict Resolution*, **32** pp. 636-670, 1988.

Saperstein, Alvin M., "Chaos—A Model for the Outbreak of War," *Nature*, **309**, pp. 303-305, 1984.

Saperstein, Alvin M., "SDI a Model for Chaos," *Bulletin of the Atomic Scientists*, pp. 24-27, Oct 1988.

Scollan, D. F., Y. Y. Azmy and V. Protopopescu, *Nonlinear Maps with Competitive Interactions: Fixed-Points, Bifurcations, and Chaotic Attractors*, ORNL/TM-11315, Oak Ridge National Laboratory, 1989.

Scott, Stephen, "Clocks and Chaos in Chemistry," *New Scientist*, pp. 53-59, 2 Dec 1989.

Sholl, D., G. D. Doolen and Y. C. Lee, *Analytic Solutions for Lanchester Equations*, Center for Nonlinear Studies, Los Alamos National Laboratories, (forthcoming).

Simon, Herbert A., "Prediction and Prescription in Systems Modeling," *Operations Research*, **38** (1), pp. 7-14, 1990.

Stewart, Ian, "Portraits of Chaos," *New Scientist*, pp. 42-47 4 Nov 1989.

Stewart, Ian, *Does God Play Dice?, The Mathematics of Chaos*, Basil Blackwell Ltd., Oxford (1989).

Thompson, J. M. T., and H. B. Stewart, *Nonlinear Dynamics and Chaos*, John Wiley, New York (1987).

Vivaldi, Franco, "An Experiment with Mathematics," *New Scientist*, pp. 46-49, 28 Oct 1989.

Walt, Stephen M., "The Search for a Science of Strategy," *International Security*, **12** (1), pp. 140-165, 1987.

Publications Subscription Rates

***PHALANX* — 4 Issues Annually**

Domestic Rates		Foreign Rates	
1-Year	$20	1-Year	$40
2-Year	$35	2-Year	$70

***Military Operations Research* — 4 Issues Annually**

Domestic Rates		Foreign Rates	
1-Year	$40	1-Year	$80
2-Year	$75	2-Year	$150

Combination Subscription — The Best Value!

Domestic Rates		Foreign Rates	
1-Year	$55	1-Year	$110
2-Year	$100	2-Year	$200

Foreign rates are based on checks drafted on US banks.

Please Start My Subscription Today!

PHALANX ❑ MOR ❑ Combination Subscription ❑

1-Year ❑ 2-Year ❑ Total Enclosed ________

Name: ______________________________

Address: ______________________________

City __________ State ______ Zip Code ________

If paying by credit card

❑ Visa
❑ Master Card
❑ American Express

____
Credit Card Number

____
Expiration Date

____
Signature

Final-Cost Estimates for Research & Development Programs Conditioned on Realized Costs

Mark A. Gallagher
Program Analysis and Evaluation
Office of the Secretary of Defense

David A. Lee
Logistics Management Institute

ABSTRACT

We apply Multiple Model Adaptive Estimation (MMAE), a proven method of system identification widely used in engineering applications, to the problem of determining Bayesian cumulative distribution functions (CDFs) of the final cost and completion time of on-going Research and Development (R&D) programs, conditioned on actual cost of work performed (ACWP) data. Modeling cumulative expenditures with Rayleigh distributions, we produce graphs of the results that give useful assessments of cost and schedule risks. The procedure is implemented in a convenient spreadsheet. We give three examples of its application to actual data, and results of a Monte Carlo analysis verify the method.

1. INTRODUCTION

Estimates of cost and duration for Research and Development (R&D) programs often increase significantly during the project. Development costs of the Concorde aircraft exceeded original estimates by more than a factor of five. In defense acquisition, where development programs for major weapon systems (aircraft, tanks, missiles) often cost billions of dollars, some development programs' final costs and completion times have been twice the original projections.

R&D programs typically undergo periodic reviews, at which estimates of the cost-to-go are critical data for decisions on whether or not to continue. At such reviews, point estimates of final cost and completion times are not particularly helpful to management because of their uncertainty. Even a firm fixed-cost development contract does not guarantee a total final cost since requests for equitable adjustment often add substantially to the costs of a program.

At an intermediate review, management needs quantitative estimates of the cost and schedule *risks* of continuing the program. They need estimates of the probability distribution of final cost and completion times, conditioned on present knowledge, for example on expenditures to date. Knowing that available information indicates that final cost and completion times are likely to fall in relatively narrow intervals or, conversely, that sets of costs and completion times occupying relatively broad intervals are all about equally likely can greatly benefit decision making.

In this paper, we develop a method for determining Bayesian probability distributions of final cost and completion times of R&D programs, from data on incurred costs (specifically, from the actual cost of work performed (ACWP) data provided in cost performance reports). A spreadsheet that is convenient for use on microcomputers implements the algorithm. With this tool, management may easily access the cost and schedule risks inherent in continuing an R&D program.

The method that we apply, Multiple Model Adaptive Estimation (MMAE) [16,17], is widely used by scientists and engineers dealing with electronic and mechanical systems. MMAE is a method for system identification, which is identifying the unknown properties of a system from observations to predict the system's future behavior. System identification is an extensively developed part of mathematical system theory. Since many tasks in cost analysis are system identification tasks, it seems helpful to apply that knowledge to them.

MMAE requires a model of the system studied, and in this paper we use the Rayleigh probability model for the time-history of expenditures in an R&D program. Several cost analysts studied the applicability of that model [1,2,6,7,11,12,20,21], and concluded that it represents R&D phases of major defense acquisition programs well.

MMAE involves the use of Kalman filters to estimate the state of a system, given noisy observations. A system's "state" is a set of parameters that describe its configuration fully, and determine its future evolution (given future inputs). For example, in Newtonian mechanics the state of a mass point is a set of three position coordinates and three velocity coordinates. In this paper, we define the state of a development project as its earned value, measured by ACWP.

The Kalman filter [13,14,15] uses a model of the system to project the Bayesian probability density of its state, conditioned on a set of noisy observations. The Kalman filter results are optimal for linear system models, Gaussian noises, and natural definitions of "optimal." The filter computations proceed iteratively and are computationally tractable.

Our application of MMAE determines the likelihood of various values of the two

Application Areas:
Cost Analysis
Research and Development

OR Methodologies:
Statistics
Kalman Filters

parameters of a Rayleigh model, based on the residuals from a set of Kalman filters. This allows us to produce graphs of the probability that the final cost or the completion time will not exceed any particular value. These graphs give managers a clear indication of the cost and schedule risk in continuing a development program.

We discuss the Rayleigh model and its applicability to R&D programs in Section 2. Section 3 presents the development of a dynamics model for earned value over time. We describe Kalman filters and MMAE as used in this application in Sections 4 and 5, respectively. In Section 6, we summarize the steps in the method. Section 7 contains sample applications, and a Monte Carlo analysis of the proposed technique is presented in Section 8. The paper concludes with a summary.

2. THE RAYLEIGH MODEL

Norden proposed that the Rayleigh distribution function can model expenditures for R&D programs [18]. He stated:

> that there are regular patterns of manpower buildup and phase-out in complex projects. ... The cycles do not depend on the nature or work content of the project but seem to be a function of the way groups of engineers and scientists tackle complex technological development problems.

Norden derived the relationship based on the assumption that the effectiveness with which problems are solved improves as a linear function of time. "Norden's description of the process is this: The rate of accomplishment is proportional to the pace of the work times the amount of work remaining to be done." [19] Putnam summarized testing of the Rayleigh model on estimating manpower for over 200 software development projects as follows:

> Many of these also exhibit the same basic manpower pattern—a rise, peaking, and exponential tail off as a function of time. Not all systems follow this pattern. ... It is because manpower is applied and controlled by management. Management may choose to apply it in a manner that is suboptimal or contrary to system requirements. [20]

Within the Department of Defense, weapon system R&D expenditures often follow a Rayleigh cumulative distribution function [1,2,6,7,11,12,20,21]. Watkins [21], Abernethy [1], Lee, Hogue and Hoffman [12], and Elrod [2] tested the ability of the Rayleigh model to fit actual weapon system R&D data. They all concluded that the Rayleigh model fits well. Lee, Hogue, and Gallagher [11] presented a procedure, based on the Rayleigh model, to determine budget profiles from an R&D estimate.

The Rayleigh model for cumulative earned value during R&D is

$$v(t) = d[1 - \exp(-\alpha t^2)] \tag{1}$$

where v represents the earned value at time t. In this paper we model earned value by expenditures (as reported by ACWP) expressed in constant dollars. The parameter d scales the Rayleigh cumulative distribution function (CDF) to costs, and the shape parameter, α, determines the time of peak rate of expenditures, t_p:

$$\alpha = \frac{1}{2t_p^2} \tag{2}$$

Since the Rayleigh distribution function has an infinite tail, the modeled expenditures would never terminate. We define the time of final development, t_f, as when 97 percent of the expenditures are complete;

$$D = v(t_f) = 0.97d \tag{3}$$

where D is the total R&D program cost. The final time relates to the time of peak rate of expenditures with $t_f = 2.65t_p$ [11]. In addition, the Rayleigh shape parameter α can be determined from a projection of the completion time with

$$\alpha = \frac{3.5}{t_f^2} \tag{4}$$

We employ the Rayleigh model to predict the change in earned value as time passes.

3. EARNED VALUE OVER TIME

A generalized model that embraces both the Rayleigh and Parr [19] models is

$$\frac{dv}{dt} = F(v) \tag{5}$$

where v is earned value and $F(v)$ gives the rate at which the project absorbs resources efficiently. The function $F(v)$ is like Parr's "number of visible jobs" to which effort can efficiently be applied. The function $F(v)$ must satisfy some common-sense conditions: $F(v)$ must be positive, except that it is zero at $v = v(t_f)$, the final value of the project, and, possibly, also at $v = v(t_0)$, the project start. $F(v)$ must be increasing in some neighborhood of $v = v(t_0)$ and decreasing in some neighborhood of $v = v(t_f)$.

If $F(v)$ is also continuous, (5) is uniquely solvable in the form

$$P(v) = t$$

where the continuously differentiable function $P(v)$ satisfies $dP/dv = 1/F$ with initial condition $P(0) = 0$. By the positivity of F, P is monotone increasing, so the inverse function P^{-1} exists, and

$$v = P^{-1}(t)$$

This formulation is a generalization of the Rayleigh case shown in (1). A straightforward calculus exercise using (1) shows that the $P(v)$ corresponding to Rayleigh is

$$F(v) = 2ad\left(1-\frac{v}{d}\right)\left[\frac{-1}{a}\ln\left(1-\frac{v}{d}\right)\right]^{\frac{1}{2}}$$

and the $F(v)$ for the Rayleigh case is

$$P(v) = \left[\frac{-1}{\alpha}\ln\left(1-\frac{v}{d}\right)\right]^{\frac{1}{2}}$$

Solving (5) with initial conditions of $v(t_i) = v_i$ for the Rayleigh case, one gets

$$\begin{aligned} v(t) &= P^{-1}(t - t_i + P(v_i)) \\ &= d\left[1-\exp\left(-\alpha\left(t-t_i+\sqrt{\frac{-1}{\alpha}\ln\left(1-\frac{v_i}{d}\right)}\right)^2\right)\right] \end{aligned} \tag{6}$$

for $t \geq t_i$. We apply the Rayleigh model as employed in (6) to predict the earned value at a future time given an earlier estimate of the earned value. Equation (6) is the dynamics model that propagates state estimates (means of the Bayesian probability distribution functions) for earned value through time in the Kalman filter formulation.

4. KALMAN FILTER

The Kalman filter is an iterative Bayesian state estimation technique. (Maybeck presents a thorough discussion in [15].) The state is the random variable of interest; in this application to R&D programs, the state is the earned value and the measurements are the reports of actual costs incurred. The first stage of the Kalman filter propagates the state distribution through time based on a dynamics model. The second stage updates the distribution with the information from an actual measurement of the system. The Kalman filter algorithm repeats these two steps for each available measurement. This section develops the propagation and update stages of a Kalman filter. In this section, we assume the three parameters required in a Kalman filter exist; in the next section, we apply Multiple Model Adaptive Estimation (MMAE) [16,17], another Bayesian technique that uses many Kalman filters each with a different combination of assumed parameters, to evaluate the likelihood of various parameter values.

For this application, we define the Kalman filter state, $x(t_i)$, as the cumulative earned value (expenditures expressed in constant dollars) at time t_i; thus, $x(t_i) = v(t_i)$. We indicate the means of Bayesian probability distributions for the Kalman filter state by a hat. At the time of each measurement, the Kalman filter algorithm calculates two state distribution means. A superscript minus sign indicates the distribution mean prior to incorporating the measurement update, $\hat{x}(t_i^-)$. Similarly, a superscript plus sign indicates the distribution mean updated with the information from a measurement at time t_i, $\hat{x}(t_i^+)$.

The steps in a Kalman filter iterate between propagation of the distribution mean through time and measurement update of the distribution mean. The state propagation is determined for the Rayleigh model in (1) with (6) as

$$\hat{x}(t_i^-) = d\left[1 - \exp\left(-\alpha\left(t_i - t_{i-1} + \sqrt{\frac{-1}{\alpha}\ \ln\left(1 - \frac{\hat{x}(t_{i-1}^+)}{d}\right)}\right)^2\right)\right] \quad (7)$$

The appropriate initial state distribution mean in this application is zero because no expenditures can be incurred before the beginning of the program; $\hat{x}(t_o) = 0$.

The measurement update step incorporates the new information from a measurement. The notation for the measurement at time t_i is z_i. In this application, the measurement is the value of ACWP reported in the cost performance reports, adjusted for inflation. Since the measurement is a direct measure of the state, the Kalman filter residual is the difference between the measurement and the mean of the state distribution prior to incorporating the measurement:

$$r_i = z_i - \hat{x}(t_i^-) \quad (8)$$

The Kalman filter gain, k, weights the information provided by the dynamics model along with the prior measurements and the information provided by the new measurement. Thus, the Kalman filter algorithm calculates the updated state distribution mean with

$$\begin{aligned} \hat{x}(t_i^+) &= \hat{x}(t_i^-) + kr_i \\ &= (1-k)\,\hat{x}(t_i^-) + k\,z_i. \end{aligned} \quad (9)$$

Since the Kalman filter gain provides the relative weighting of two pieces of information about the system available at the time, the gain is bounded between zero and one; $0 \le k \le 1$. If the gain is zero, the update distribution mean is based entirely on the dynamics model; whereas if the gain is one, the updated mean is the last measurement. With values for d, α , and k, one can apply a Kalman filter using (7), (8) and (9) iteratively for each available measurement (reported actual cost). The next section presents a development for Bayesian estimation of these three parameters.

5. MULTIPLE MODEL ADAPTIVE ESTIMATION (MMAE)

MMAE is a Bayesian system identification technique that estimates unknown system parameters when applying Kalman filters [16,17]. In this application, we use MMAE to determine the likelihood of parameters d (cost scale parameter), α (Rayleigh shape parameter), and k (Kalman filter gain). The advantage of applying MMAE is that the probabilities are conditional on the actual cost data, which prevents assigning probabilities to final costs below the incurred cost or completion times less than the elapsed duration.

An overview of the algorithm follows: The set-up for employing MMAE is to discretize the continuous space for each parameter into a set of representative points. The MMAE algorithm processes the measurements (reported actual costs in this application) through a Kalman filter at each combination of discrete parameters. Each filter's residuals determine the probability of that filter's parameters being correct, conditioned on the measurements processed to that time. After processing all the available measurements, the filter probabilities indicate the likelihood of the parameters in that filter being correct conditioned on the measurements. We relate the filter parameter d to total program cost with (3) and the filter parameter α to project duration with the relationship in (4). Thus, after processing all available data, each final filter probability represents the likelihood of that filter's corresponding completion cost, time and Kalman filter gain conditioned on the actual cost reports.

We convert these final filter probabilities to cumulative distribution functions (CDFs) conditioned on the cost reports for either the final cost or completion time. The cumulative probability that the final cost is less than any particular value is determined by summing all the filter probabilities with corresponding final costs equal to or less than that value. We sort in increasing order the final cost values associated with the filters along with their final probabilities. We generate a CDF by incrementally summing the final filter probabilities as the filter parameters for d increase. Similarly, we incrementally sum the filter probabilities as the values for α increase to determine a CDF for project duration.

The details of the algorithm begin with the set-up for applying MMAE, discretizing the parameter space. Define the number of Kalman filters as L. Let $\boldsymbol{a}_l$ represent the vector of parame-

ters d_l, a_l, and k_l selected for the lth filter, where $l = 1, ..., L$. With a vector of parameters $\boldsymbol{a}_l$, one can process the data through a Kalman filter by iteratively applying (7), (8) and (9). In the examples and Monte Carlo analysis, we used 20 values for d, 20 values for α, and 5 values for k, equally spaced in each dimension. Thus, we processed the reported cost data through 2,000 Kalman filters.

Our approach discretized the parameter space in two steps. The first step is processing the currently available data through filters with a coarse discretization, and the second step is refining the discretization based on the filters' sum of squared residuals. Let the measurement history be represented by $\boldsymbol{Z}_N = \{z_1, z_2, \ldots, z_N\}$ where z_i is the cumulative cost incurred at time index t_i. We determine the range of the Rayleigh parameter α from estimates of the minimum and maximum completion time with (4), and we varied the values of α incrementally over the range. The default range for estimated completion times is from a minimum of the last cost report, t_N, to an arbitrary maximum time of 15 years. For example, if the maximum completion time is

15 years, $\alpha_{\max} = \frac{3.5}{15^2} = 0.156$ from (4). ($\alpha_{\max}$ is actually the smallest shape parameter.) Our algorithm sets the minimum value for the cost scale parameter equal to the last reported cost, $d_1 = z_N$ and sets the maximum value equal with the amount and time of the last cost report with the Rayleigh curve for the longest program,

$$d_m = \frac{z_N}{1 - \exp(-\alpha_{\max} t_N^2)} \tag{10}$$

The Kalman filter gain range is $0 \le k \le 1$. An analyst may adjust either the cost parameter or completion time ranges. The algorithm processes the cost data through each of the Kalman filters with this initial coarse discretization of the parameter space. We use this first pass through the data to estimate the residual variance and to refine the parameter discretization.

MMAE determines the filters' probabilities by the magnitude of that filter's residuals. The Kalman filter residuals for linear systems with known structural matrices and driven by white noise are independent and Gaussian distributed with zero mean and known variance [15]. Although these assumptions are not met in this case, other applications assumed that the residuals are Gaussian and obtained useful results [3,4,5,8,9]. We assumed that the residuals calculated with (8) are zero mean with a variance estimated from the Kalman filter with the smallest sum of squared residuals from an initial pass through the data;

$$s_r^2 = \min\left[\frac{1}{N-1} \sum_{i=1}^{N} (r_i^l)^2 \;\middle|\; l = 1, \ldots, L\right] \tag{11}$$

After the first pass of the data through the bank of Kalman filters, we reduce the parameter range to eliminate parameter values that resulted in sum of squared residuals greater than three times the minimum value, s_r^2. Our algorithm equally spaces the parameters for the Kalman filters across the reduced parameter ranges. The algorithm calculates the MMAE probabilities on the second pass of the data through the Kalman filters. Based on the assumption of zero mean and the residual variance estimated in (11), the Gaussian probability density function for the ith measurement, z_i, conditioned on the lth filter's vector of parameters, $\boldsymbol{a}_l$, and the prior measurement history, $\boldsymbol{Z}_{i-1}$, is

$$f(z_i \mid \boldsymbol{a}_l, \boldsymbol{Z}_{i-1}) = \frac{1}{\sqrt{2\pi\, s_r^2}} \exp\left(\frac{-r_i^2}{2 s_r^2}\right)$$

as adapted from Equation (10-98) in Reference [16]. The probability for the jth filter having the "correct" parameters conditioned on the measurement history through time t_i is

$$p_j(t_i \mid \boldsymbol{Z}_i) = \frac{f(z_i \mid \boldsymbol{a}_j, \boldsymbol{Z}_{i-1})\, p_j(t_{i-1} \mid \boldsymbol{Z}_{i-1})}{\sum_{l=1}^{L} f(z_i \mid \boldsymbol{a}_l, \boldsymbol{Z}_{i-1})\, p_l(t_{i-1} \mid \boldsymbol{Z}_{i-1})} \tag{12}$$

from Equation (10-104) in Reference [16]. The probabilities at each measurement time, t_i for $i = 1, ..., N$, must sum to one;

$$\sum_{l=1}^{L} p_l(t_i \mid \boldsymbol{Z}_i) = 1. \tag{13}$$

This normalization limits the conditional probabilities to only the L discrete parameter combinations used in the filters.

The initial or *a priori* probabilities account for information available about the likelihood of particular filter combinations before the measurement data are processed. If no information is available, the *a priori* probabilities should all be equal; $p_l(t_0) = 1/L$ for $l = 1, ..., L$. In addition, if any of the filter probabilities became zero, that filter's probabilities, calculated with (12), would remain zero for all the later times. To prevent prematurely discarding potentially viable filter parameters, practitioners commonly apply a heuristic [16,17]; if any of the filter probabilities decreases below a very small lower bound, such as 0.0001, the heuristic artificially increases that filter's probability to the lower bound. The filter probabilities that result after the last datum are not adjusted with this heuristic. The final filter probabilities represent the likelihood of each combination of model parameters conditioned on the available measurement history, $\mathbf{Z}_N$.

We use the filter probabilities to determine estimates for the final cost and completion time. The final cost corresponding to the parameters d_l and α_l is $D_l = d_l T_{\alpha_l}$; the translation factor, T_{α_l}, is 0.97 from (3) for D_l expressed in constant dollars. To express D_l in current dollars, the translation factor must account for inflation during the program. Let the sequence of start of fiscal years during the program duration be represented by $\tilde{t}_i$ where $\tilde{t}_0$ is the program start and $\tilde{t}_{\tilde{N}}$ is the projected program end. Further, let the corresponding inflation indices for the following fiscal year be $\tilde{r}_i$. Then the translation factor corresponding to α_l is $T_{\alpha_l} = \sum_{i=0}^{\tilde{N}-1} \tilde{r}_i (e^{-\alpha \tilde{t}_i^2} - e^{-\alpha \tilde{t}_{i+1}^2})$. Each D_l should be constrained to be greater than or equal to last cost report, expressed appropriately in constant or current dollars. The mean estimate of final cost conditioned on the available measurement history, $\mathbf{Z}_N$, is calculated with

$$\hat{D} = \sum_{l=1}^{L} D_l \, p_l(t_N \mid \mathbf{Z}_N) \tag{14}$$

where t_N is the time index corresponding to the last available cost report, Z_N. Similarly, the conditional mean estimate for completion time, based on the (4), is

$$\hat{t}_f = \sum_{l=1}^{L} \left(\frac{3.5}{\alpha_l} \right)^{0.5} p_l(t_N | \mathbf{Z}_N) \tag{15}$$

The estimates from (14) and (15) are the MMAE means conditioned on the actual cost data.

The cost CDF conditioned on the measurement history shows the probability that the final cost, D, will be less than any dollar value. Let the cost scale parameters increase from d_1 to d_m. The sum of filter probabilities for all the filters with d_i represents the probability over the range $[0.5(d_{i-1} + d_i),\ 0.5(d_i + d_{i+1})]$ for $i = 2, ..., m\text{-}1$. There is no conditional probability below d_1 or above d_m . Define $\tilde{d}_0 = d_1$, $\tilde{d}_i = 0.5(d_i + d_{i+1})$ for $i = 1, \ldots, m\text{-}1$, and $\tilde{d}_m = d_m$. The final cost estimate for the Rayleigh model with parameters d_l and α_l is calculated as $\tilde{D}_l = T_{\alpha_l} \tilde{d}_l$ where T_{α_l} is the translator factor to constant or current dollars used for (14). All final cost estimates should exceed the last datum. We calculate the cumulative probabilities by summing the filter probabilities for filters with final cost estimates, $\tilde{D}_l$, less than a dummy cost variable λ with linear interpolation between cost estimates; with $\tilde{D}_l$ for $l = 0$ to L sorted in increasing magnitude:

$$P(D < \lambda | \mathbf{Z}_N) = \begin{cases} 0 & \text{if } \lambda \le \tilde{D}_0 \\ \dfrac{\lambda - \tilde{D}_1}{\tilde{D}_2 - \tilde{D}_1} p_1(t_N | \mathbf{Z}_N) & \text{if } \tilde{D}_0 \le \lambda \le \tilde{D}_1 \\ \sum_{j=1}^{l-1} p_j(t_N | \mathbf{Z}_N) + \dfrac{\lambda - \tilde{D}_l}{\tilde{D}_{l+1} - \tilde{D}_l} p_l(t_N | \mathbf{Z}_N) & \text{if } \tilde{D}_l \le \lambda \le \tilde{D}_{l+1} \\ & \text{for } 1 \le l \le L-1 \\ 1 & \text{if } \lambda \ge \tilde{D}_L \end{cases} \tag{16}$$

As a simple example, suppose we applied this approach with L=5, d_l =100, 120, 140, 160, and 180, and corresponding final probabilities $P_l(t_N | \mathbf{Z}_N)$ = 0.1, 0.2, 0.5, 0.1, and 0.1. Table 1 shows the calculations in the columns with the index in the first column. The values of d_l are parameters used in the filters. The values of $\tilde{d}_l$ are the upper end of the range where d_l is the nearest d_i and restricts the values at each end. The calculations for each $\tilde{D}_l$ adjust for program duration and inflation as shown for (14). ($T_{\alpha_l} = 0.97$ for constant dollar values.) The next to last column contains the final filter probabilities, and the last column sums these probabilities as the cost estimates $\tilde{D}_l$ increase. The cumulative probability at $\tilde{D}_l$, $P(D < \tilde{D}_l | \mathbf{Z}_N)$equals $\sum_{i=1}^{l} P_i(t_N | \mathbf{Z}_N)$.

Table 1. Sample Cost Cumulative Distribution Function (CDF) Calculations

l	d_l	$\tilde{d}_l$	$\tilde{D}_l$	$P_l(t_N \mid \boldsymbol{Z}_N)$	$P(D < \tilde{D}_l \mid \boldsymbol{Z}_N)$
0		100	97.0		0.0
1	100	110	106.7	0.1	0.1
2	120	130	126.1	0.2	0.3
3	140	150	145.5	0.5	0.8
4	160	170	164.9	0.1	0.9
5	180	180	174.6	0.1	1.0

The cost CDF is a linear interpolation between the pairs of points $\tilde{D}_l$ and $P(D < \tilde{D}_l \mid \boldsymbol{Z}_N)$.

The graph of $P(D<\lambda \mid \boldsymbol{Z}_N)$ versus λ shows the conditional probability that the final cost will be less than the any particular value. Finer discretization smoothes the CDF graph. The constant-dollar and current-dollar curves have very similar shapes. We also generate the CDF for project duration—equivalently completion time— using the parameter α and the relationship in (4).

A confounding relationship limits the ability to estimate both d and α when αt^2 is small [12]. This problem can be seen by expanding the exponential in (1);

$$v(t) = d\,(1 - \exp(-\alpha\, t^2)$$

$$= d\left[1 - (1 - \alpha\, t^2 + \frac{\alpha^2 t^4}{2} \ldots)\right] \approx \alpha\, d\, t^2 + O(d\,\alpha^2 t^4)$$

where the function $O(\cdot)$ represents higher order terms. When αt^2 is small, the higher order terms are negligible and only the product of α and d, but not their individual values, can be estimated from the data. The relationship $\alpha t^2 < 0.5$ holds prior to the time of peak expenditure rate, as seen from (2). Thus, many different Rayleigh curves appear to fit the data from t_0 to t_p due to the canceling effects of changes to α and d. With an independent estimate for either the time of peak rate of expenditures or the completion time, an analyst may determine the parameter α and estimate d using the data. MMAE has the same confounding problem as any statistical technique when only data before the peak expenditure rate is available. If an independent estimate of α is available, one can put that value into all the filters and apply MMAE to estimate the probability distribution of the final cost.

6. ALGORITHM STEPS

While the development of the algorithm is complex, implementation is not difficult. An Excel spreadsheet with a Visual Basic Module that applies this technique is available from the authors. The runtime on a 486 computer is about 1 minute with 50 data points. The procedure steps are enumerated below:

Step 1) Adjust the history of cost reports for inflation
- Determine the delta between cumulative cost reports
- Apply the appropriate inflation index to the delta
- Sum the constant dollar deltas to obtain cumulative costs in constant dollars
- Determine time indices in years for each datum from the program start date

Step 2) Determine the completion time range (may be fixed to a single value)
- Default range is from the time index of the last cost report to 15 years (arbitrary)
- Adjust completion time range based on program knowledge
- Relate completion time range to corresponding α range with (4)

Step 3) Determine the range for final cost estimates
- Default for minimum value is last reported incurred costs (in constant dollars)
- Default for maximum value is estimated with (10)
- Adjust final cost range based on program knowledge
- Relate final cost range to range for d with (3)

Step 4) Initialize Kalman filters
- Set number of discrete points for each variable, such as 20 for d and α with 5 for k
- Determine discrete values equally spaced across selected parameter range
- Assign variables for a Kalman filter with each combination of parameter values
- Set prior mean of state distributions to zero at initial time index, t_0; $\hat{x}(t_0) = 0$

Step 5) Process available data through filters to estimate residual variance and adjust parameter ranges
- Propagate state distribution means

with (7)

- Update state distribution means with (8) and (9)
- Collect sum of squared residuals from (8) for each Kalman filter
- Find minimum sum of squared residuals and estimate residual variance, s_r^2, with (11)
- Reduce α and d ranges to eliminate values that always resulted in sum of squared residuals greater than 3 times the minimum sum
- Equally space the filter parameters across the reduced parameter ranges
- Reset prior means and set filter probabilities $p_l(t_0) = 1/L$ for $l = 1, ..., L$.

Step 6) Process data values through bank of filters to determine filter probabilities

- Propagate state distribution means with (7)
- Update state distribution means with (8) and (9)
- Calculate filter probabilities with (12)
- Normalize filter probabilities to meet (13)
- Except for last data point, adjust filter probabilities for lower bound; $p_l(t_i) \geq 0.0001$

Step 7) Determine conditional probabilistic-weighted averages with (14) and (15)

Step 8) Determine cost conditional CDF with (16)

7. SAMPLE APPLICATIONS

We applied the Bayesian estimation

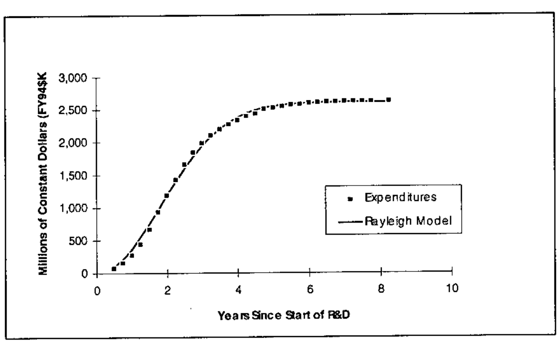


Figure 1. F-15 Airframe Contract Expenditures

approach to three diverse historic programs, the F-15 airframe development, the NavStar Global Positioning System (GPS) Satellite, and the MK 50 Torpedo. We selected these programs to cover a variety of technologies without prior knowledge of how well the Rayleigh model fits them. The F-15 development contract completed on schedule with very slight cost growth. The satellite program experienced much higher final cost than originally projected. The MK 50 program required a substantial schedule increase beyond the originally projected development time and almost twice the expense of original cost estimate. The only program data used in our approach were the originally projected duration and the actual cost reports. We set the completion time ranges from the originally projected length to twice that length in each application. The F-15 airframe development contract started in January 1970. The contract continued for over 8 years, but most of the earned value occurred in the first 5 years. The Rayleigh model fits the reported expenditures reasonably. Figure 1 shows the Rayleigh model with the least squares

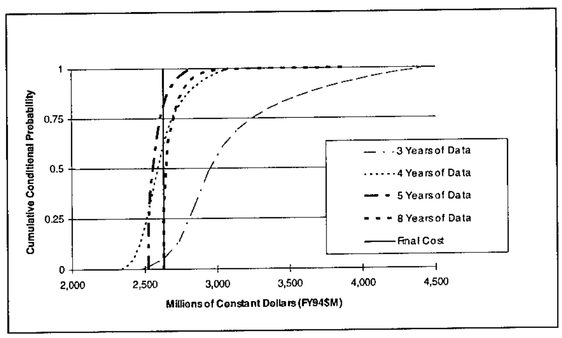


Figure 2. F-15 Airframe Development Final Cost CDFs

parameters and the cost reports adjusted for inflation.

We applied our Bayesian cost estimation approach with the initial 3, 4, 5, and all years of F-15 airframe expenditures. We set the completion time range from 5 to 10 years. Figure 2 depicts the resulting CDFs. The CDF based on only 3 years of data indicates a wide potential range for the final cost. When 4 or more years of data were used, the CDFs are very close to the actual final cost.

Most of the techniques used today to predict final costs of R&D programs give a point estimate for the final cost. Sophisticated decision makers want more information than that, of course. They may well find the MMAE method's risk assessments that provide the probability

that the final cost or program duration is contained in any given range quite helpful.
Nevertheless, to compare with other techniques, we had to select a point estimate. We compared the MMAE probabilistic-weighted average value calculated using (14) with four commonly applied techniques that predict point estimates for final program cost [10]. For each of the four techniques, the final cost estimate is actual cost of work performed (ACWP) plus the quotient of work remaining divided by a cost performance index (CPI). Work remaining is budgeted work minus management reserve and budgeted cost of work performed (BCWP). The four techniques vary in the calculation of the CPI. The index for cumulative CPI (Cum CPI) is the cumulative BCWP performed divided by the cumulative ACWP. Similarly, CPI-3 and CPI-6 are calculated with BCWP and ACWP over the last 3 or 6 months, respectively. The cumulative CPI times cumulative schedule performance index (CPI*SPI) is the CPI multiplied by the cumulative budget cost of work scheduled divided by the ACWP. We compare these four techniques with the MMAE average from (14) for the three historical programs. These four techniques depend upon the accuracy of the program baseline, whereas the MMAE approach has the advantage of being independent of the projected program budget.

Table 2 depicts the various final cost estimates for the F-15 airframe development. The CPI techniques were low with the initial cost data and increased over time. In contrast, the probabilistic means from the Rayleigh/MMAE approach started much too high with only data prior to peak expenditure rate and decreased with additional data.

The second sample program is NavStar Global Positioning System (GPS) Satellite. This R&D program, which began in June of 1974, had a projected completion time of 4.3 years and a projected final cost of 40 million in current dollars. The program required almost 6 years and required 116.3 million in current dollars. The cumulative costs in constant dollars appear to fit the Rayleigh model; Figure 3 depicts the rate of expenditures and the derivative of the least squares Rayleigh model. We calculated the expenditure rates as the increase in reported cumulative expenditures divided by the time delta between cost reports. The Kalman filter gain, k, accounts for the measurement noise in the cost reports, apparent from the variation in reported expenditure rates. A quick heuristic, based on the Rayleigh model, to evaluate progress in R&D programs is that 60 percent of the expenditures occur after the time of peak expenditure rate, t_p.

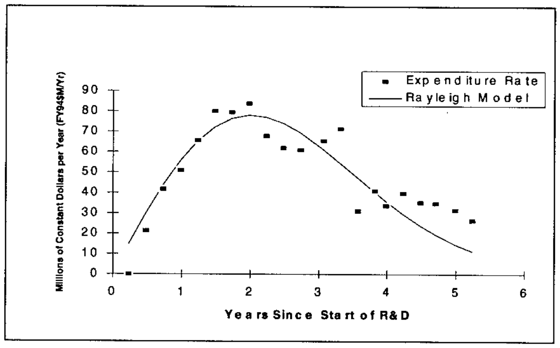


Figure 3. NavStar GPS Satellite Rate of Expenditures

We applied the Bayesian method with 2, 3, 4, and 5 years of expenditure data, and Figure 4 depicts the final cost CDFs. Without data after the peak rate of expenditure time, the completion time and final cost are statistically confounded. Since the peak rate occurs just after 2 years in the NavStar R&D, the CDF based on 2 years of data indicates the potential for a very long and expensive program. The level expenditure rate during the fourth year, shown in Figure

Table 2. F-15 Airframe Contract Final Cost Estimates (Current Dollars)

Years of Data	CUM CPI	CPI-3	CPI-6	CPI*SPI	Rayleigh/MMAE
2	779.9	752.4	764.8	784.7	1,880.4
3	775.9	779.4	775.3	777.3	1,016.3
4	678.4	696.8	689.4	681.2	834.6
5	815.6	820.9	819.0	816.2	836.3

The program manager estimate in Mar 1978 (8.25 years) was 850.0.

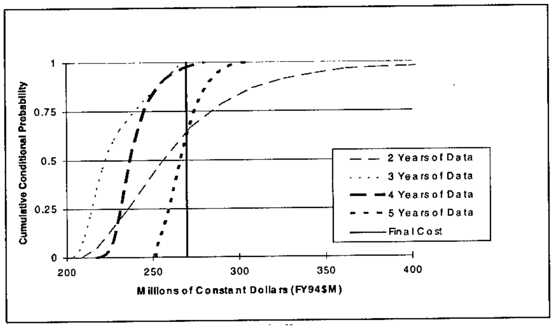


Figure 4. NavStar GPS Satellite Final Cost CDFs

3, resulted in the CDFs based on 3 and 4 years of data to underestimate the final cost. With 5 years of data, the CDF is very accurate.

We used the final filter probabilities from the same runs to generate the program duration CDFs. The duration range was from the original projection of 4.3 years to 8.6 years. The duration CDFs, shown in Figure 5, remain fairly consistent until 5 years of data was used. Figure 3 shows that the fourth year of data had a higher rate of expenditures than predicted with the Rayleigh model; the CDF conditioned on 5 years of data indicate an increased probability of a longer program. Data fluctuations seems to affect the program duration CDFs more than the functions for final cost.

We present the various final cost estimates in Table 3. The CPI techniques were low initially and increased with more data. In contrast, the MMAE averages remained slightly below the actual final cost.

The final example is the development of the MK 50 Torpedo. This program began in August of 1983 with a 5 year projected duration. The program was extended an additional 3 years, and the final costs increased 65 percent higher in current dollars. The completion time range was set from 5 to 10 years. The CDFs are depicted in

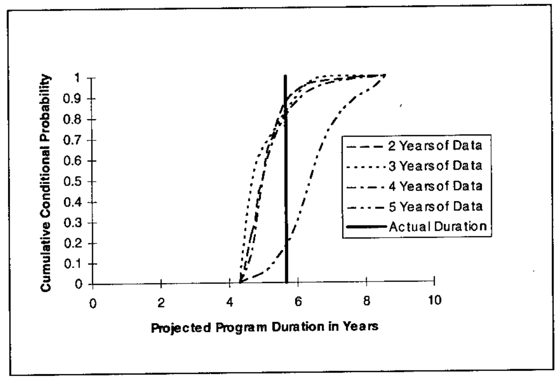


Figure 5. NavStar GPS Satellite Duration CDFs

Figure 6. A small probability of the cost being as high as the actual final cost is seen with even 3 years of data. With each year of additional data, the median value from the CDFs moves closer to the actual final cost. With 6 and 7 years of data, much of the CDFs exceed the final cost because the lower bound of the curves is cost incurred to that point in time.

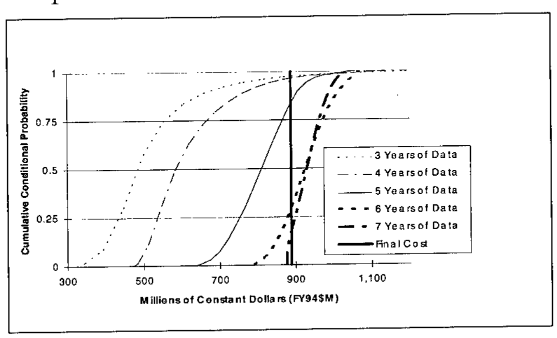


Figure 6. MK 50 Torpedo Final Cost Likelihood Curves

The various final-cost point estimates, shown in Table 4, increased significantly. All the techniques started too low an increased as additional data was available.

These three examples demonstrate the capa-

Table 3. NavStar GPS Satellite Final Cost Estimates (Current Dollars)

Years of Data	CUM CPI	CPI-3	CPI-6	CPI*SPI	Rayleigh/MMAE
2	70.0	80.9	78.7	71.0	109.8
3	99.8	100.3	101.7	103.3	96.2
4	104.4	108.7	104.3	106.9	98.6
5	114.0	114.5	114.4	115.5	112.2

The program manager estimate in Aug 1979 (5.25 years) was 116.3.

bilities of this Bayesian cost estimation approach for on-going R&D programs. In each of the applications, the algorithm made final cost CDFs that are very near the actual final costs based on very little program specific data. Tighter bounds on the possible range of final cost or completion times based on additional program knowledge would improve the proposed method's results. A Monte Carlo analysis, presented in the next section, shows the statistical effectiveness of this approach.

8. MONTE CARLO ANALYSIS

We conducted a Monte Carlo analysis of this technique with noise-corrupted Rayleigh data to verify its statistical validity. We evaluated the algorithm estimates for accuracy of point estimates and accuracy of the final cost CDFs. The performance statistics were collected after applying the algorithm with various amounts of the generated data.

Various final costs, completion times, and noise levels determined specific cases. We generated cumulative cost reports for each fiscal quarter. The data reflected actual data in that the initial cost at time zero was zero, the cumulative cost always increased, and the cost at completion time was the final cost. For each cost report, the generated datum was calculated with

$$z_i = v(t_i) = d\left[F(t_{i-1}) + \left(F(t_i) - F(t_{i-1})\right)(1+\varepsilon)\right]$$

such that $v(t_0) = 0$, $v(t_f) = D$, and $v(t_i) \geq v(t_{i-1})$ where, $F(t) = 1 - \exp(-\alpha\, t^2)$ the Rayleigh CDF, d is from (3), α is from (4), and ε is a uniform random variable between plus and minus the noise level.

We tested seven cases. The final costs used were 2,000, 1,500 and 1,000 for a 12 year program. The Rayleigh shape parameters were determined with (4), and the noise level for 5 cases was set at 0.1. We varied the noise level to 0.2 and 0.3 for the 12 year program with final cost of 1,000 dollars. We also varied the completion time for the 1,000 dollar program to 9 and to 6 years. For each case, summary statistics were collected across 500 data sets; we applied the algorithm both with and without using the known completion times. In all the tables to be presented, the first three columns define the case by giving the true final cost, true program completion time and the noise level used to generate the data. The next sets of columns show results based on increasing amounts of data used in the estimates. For example, the column with "Time of Estimate" of 3 indicates that 3 years of quarterly data were used to calculate the statistics in that column. We define errors as the estimated value minus the true value. The top halves of the tables are results based on estimated completion times, and the bottom halves present the results when the program completion time is known, in essence estimated perfectly.

The first measure of effectiveness is the accuracy of the MMAE probabilistic mean in estimating the true cost used to generate the data. We calculated the probabilistic mean with (15) and adjusted to final cost with (3). Table 5 shows the statistics for the seven cases. For a 12 year program with unknown completion time, the results with 3 years of data have large errors and corresponding large standard deviations. This is a result of the statistical indeterminacy between the cost scale parameter and the Rayleigh shape parameter. If the final time of the program is known, the errors in the final costs are very small as seen in the bottom half of Table 5. The errors with unknown completion times are conservative in that they estimate the program to be

Table 4. MK 50 Torpedo Final Cost Estimates (Current Dollars)

Years of Data	CUM CPI	CPI-3	CPI-6	CPI*SPI	Rayleigh/MMAE
3	580.7		566.3	589.0	409.8
4	529.8	540.1	536.8	527.8	559.8
5	655.7	667.1	659.4	650.6	629.4
6	685.3	678.8	680.8	685.0	746.3
7	707.3	706.9	706.6	708.9	714.5

The program manager estimate in Dec 1990 (7.25 years) was 711.4.

much higher in cost and longer than it actually was. With data that encompass half the actual completion time, the errors become very small in comparison with the final cost with relatively small variance.

The first three cases show the linear effect for changes in the true final cost. Since we used the same seed in the random number generator, the error statistics are exactly proportional to the true final cost. The third through fifth cases show that as the noise levels increase so do the estimate standard deviations. The errors for the shorter programs in the last two cases are less because proportionately more program data was used for the estimates. In all cases, the error statistics improve with additional data, and the errors are very small when the completion time was known. We did not include the results for the median because of their similarity.

The second statistics depict the effectiveness in quantifying the cost risk of continuing a program. The cost risk is depicted with the cost CDFs generated with (16). We evaluated the CDFs by collecting the frequency with which the true cost was less than the predicted 30th and 70th percentiles. Table 6 shows the statistics for 500 runs for each case. When the reported frequency for the 30th percentile exceeds 0.30, the CDF estimates were too high. Following the trend of the mean and median, the 30th percentile was high initially and decreased as the amount of data increased. When all the data was used, the entire cost CDF exceeds the value of the last data point, which was the true final cost, because the cumulative cost projects always exceed reported incurred costs. When the final time was known, the 30th percentiles were slightly low and the 70th percentiles were slightly high.

The final measure of effectiveness is the width of the 40 percent probability interval that could be formed from the 30th to the 70th percentiles. The probability interval widths indicate the accuracy the algorithm assigns to mean estimates in Table 5. Table 6 shows that these assigned accuracies are commensurate with their true accuracies. Table 7 shows that as the additional data was used in the algorithm the probability interval widths become very small. The point estimator error with all the data was less than 0.2 percent of the true final cost, and the corresponding 40 percent probability interval width was less than 92 percent of the true final cost.

Table 5. Probabilistic Mean Estimator Statistics

	Case		Average Errors				Error Standard Deviations			
Final	Final	Noise	Time of Estimate				Time of Estimate			
Cost	Time	Level	3	6	9	12	3	6	9	12
Estimated Final Cost and Estimated Final Time										
2,000	12	0.1	434.0	5.7	1.3	0.3	562.6	13.1	2.0	0.3
1,500	12	0.1	325.6	4.3	0.9	0.2	422.1	9.8	1.5	0.2
1,000	12	0.1	217.1	2.9	0.6	0.1	281.4	6.6	1.0	0.2
1,000	12	0.2	208.9	3.2	0.0	0.7	302.6	18.9	3.0	0.6
1,000	12	0.3	161.9	8.2	0.1	1.0	312.7	44.2	5.3	0.9
1,000	9	0.1	70.1	3.9	0.2		186.6	9.9	0.3	
1,000	6	0.1	11.8	1.8			31.8	0.5		
Estimated Final Cost with Given Final Time										
2,000	12	0.1	-0.3	-0.1	0.0	0.1	20.7	4.5	1.2	0.2
1,500	12	0.1	-0.2	-0.1	0.0	0.1	15.5	3.4	0.9	0.2
1,000	12	0.1	-0.2	-0.1	-0.1	0.0	10.4	2.3	0.6	0.1
1,000	12	0.2	0.4	0.0	0.2	0.4	20.3	5.0	1.4	0.4
1,000	12	0.3	3.3	-0.1	-0.1	0.5	31.0	7.6	1.9	0.5
1,000	9	0.1	-1.1	0.2	0.1		9.0	2.2	0.2	
1,000	6	0.1	-8.2	0.3			6.1	0.3		

9. SUMMARY

We developed and tested a method of estimating the probability of final cost and completion time for R&D programs conditioned on actual cost reports. The method is based on assuming that the cumulative earned value (represented by constant-dollar expenditures) of the

Table 6. Estimated Percentile Efficiencies

Case			Frequency < 30th Percentile				Frequency ≤ 70th Percentile			
Final Cost	Final Time	Noise Level	Time of Estimate 3	6	9	12	Time of Estimate 3	6	9	12
Estimated Final Cost and Estimated Final Time										
2,000	12	0.1	0.54	0.38	0.26	1.00	0.75	0.96	0.98	1.00
1,500	12	0.1	0.54	0.38	0.26	1.00	0.75	0.96	0.98	1.00
1,000	12	0.1	0.54	0.38	0.26	1.00	0.75	0.96	0.98	1.00
1,000	12	0.2	0.57	0.31	0.10	1.00	0.79	0.95	0.83	1.00
1,000	12	0.3	0.50	0.37	0.13	1.00	0.75	0.82	0.85	1.00
1,000	9	0.1	0.26	0.48	1.00		0.71	0.87	1.00	
1,000	6	0.1	0.47	1.00			0.90	1.00		
Estimated Final Cost with Given Final Time										
2,000	12	0.1	0.23	0.17	0.11	1.00	0.77	0.82	0.92	1.00
1,500	12	0.1	0.23	0.17	0.11	1.00	0.77	0.82	0.92	1.00
1,000	12	0.1	0.23	0.17	0.11	1.00	0.77	0.82	0.92	1.00
1,000	12	0.2	0.29	0.22	0.17	1.00	0.69	0.79	0.85	1.00
1,000	12	0.3	0.32	0.21	0.10	1.00	0.72	0.78	0.86	1.00
1,000	9	0.1	0.43	0.15	0.98		0.47	0.88	1.00	
1,000	6	0.1	0.10	0.99			0.10	1.00		

Note: The theoretical standard deviation of these frequencies is 0.0205.

Table 7. Estimated 40 Percent Probability Interval Width

Case			Probability Interval Width (Distance Between 30th and 70th Percentiles)			
Final Cost	Final Time	Noise Level	Time of Estimate 3	6	9	12
Estimated Final Cost and Estimated Final Time						
2,000	12	0.1	343.5	22.9	7.1	1.0
1,500	12	0.1	257.4	17.2	5.4	0.8
1,000	12	0.1	171.6	11.5	3.6	0.5
1,000	12	0.2	197.7	20.9	7.8	1.2
1,000	12	0.3	226.4	48.2	10.9	1.6
1,000	9	0.1	197.3	16.5	1.0	
1,000	6	0.1	33.9	3.4		
Estimated Final Cost with Given Final Time						
2,000	12	0.1	25.1	7.7	4.0	0.7
1,500	12	0.1	18.8	5.8	3.0	0.8
1,000	12	0.1	12.6	3.8	2.0	0.4
1,000	12	0.2	20.7	7.6	3.5	1.0
1,000	12	0.3	29.1	11.3	5.4	1.2
1,000	9	0.1	1.5	2.8	1.0	
1,000	6	0.1	0.0	2.5		

development program followed a Rayleigh distribution. The approach uses Multiple Model Adaptive Estimation (MMAE), which employs a large number of Kalman filters, to estimate the Rayleigh model parameters. The MMAE technique, as applied in this application, provides the probabilities of various final cost estimates and projected completion times conditioned on actual cost data. We summed those probabilities to produce final cost CDFs. These CDFs depict the probability that the final cost estimate will be below various cost estimates. The final cost estimates and CDFs can be converted from constant dollars to current dollars. Similarly, CDFs for completion time can be constructed.

REFERENCES

[1] Abernethy, T., *An Application of the Rayleigh Distribution to Contract Cost Data*, Master's Thesis, Naval Postgraduate School, Monterey, California, 1984.

[2] Elrod, S. M., *Engineering and Manufacturing Development Cost Estimation: An Analysis of Combined Time-Phased and Classical Techniques*, Master's Thesis, Naval Postgraduate School, Monterey, California, June, 1993.

[3] Gallagher, M. A., *Identification of the Initial Transient in Discrete-Event Simulation Output Using the Kalman Filter*. PhD Dissertation, Department of Operational Sciences, Air Force Institute of Technology, Wright-Patterson AFB, Ohio, September, 1992.

[4] Gallagher, M. A., Bauer, K. W., Maybeck, P. S., "Initial Data Truncation for Univariate Output of Discrete-Event Simulations Using the Kalman Filter," *Management Science*, Vol. 42, 1996, pp. 559-575.

[5] Gallagher, M. A., Bauer, K. W., Maybeck, P. S., "Initial Data Truncation for Multivariate Output of Discrete-Event Simulations Using the Kalman Filter," *Annals of Operations Research*, Vol. 53, 1994, pp. 419-441.

[6] Harmon, B. R., Ward, L. M., Palmer, P. R., "Schedule Estimating Relationships for Tactical Aircraft," T. R. Gulledge, L. A. Litteral (Editors), *Lecture Notes in Economics and Mathematical Systems*, Springer-Verlag, 1989, pp. 259-280.

[7] Hogue, Michael R., "Rayleigh Curve Fitting to R&D Cost Profiles", unpublished report, OASD/PA&E, The Pentagon, June, 1990.

[8] Howard R. B., Gallagher, M. A., Bauer, K. W., Maybeck, P. S., "Confidence Intervals for Univariate Discrete-Event Simulation Output Using the Kalman Filter," in J. Swain and D. Goldsman, eds., proceedings of the 1992 Winter Simulation Conference, Arlington, Virginia, December, 1992, pp. 586-593.

[9] Howard R. B., Gallagher, M. A., Bauer, K. W., Maybeck, P. S., "Confidence Intervals Estimation for Univariate Output of Discrete-Event Simulations Using the Kalman Filter," Technical paper, Department of Operational Sciences, Air Force Institute of Technology, Wright-Patterson AFB, Ohio, October, 1992.

[10] Christen, David, Richard C. Altrolini, and John, W. McKinney, "Review of Estimate at Completion Research," *Journal of Cost Analysis*, Spring, 1995, pp. 41-62.

[11] Lee, D. A., Hogue, M. R., Gallagher, M. A., "Determining a Budget Profile from a R&D Cost Estimate," to appear *Journal of Cost Analysis*.

[12] Lee, D., Hogue, M., Hoffman, D., *Time Histories of Expenditures for Defense Acquisition Programs in the Development Phase*, presented at the 1993 Annual Meeting of the International Society for Parametric Analysis.

[13] Kalman, R.E., "A New Approach to Linear Filtering and Prediction Problems," *Transaction of ASME Journal of Basic Engineering*, Vol. 82, 1960, pp. 34-45.

[14] Kalman, R.E., Bucy, R.S., "New Results in Linear Filtering and Prediction Theory," *Transaction of ASME Journal of Basic Engineering*, Vol. 83, 1961, pp. 95-108.

[15] Maybeck, P. S., *Stochastic Models, Estimation, and Control: Volume 1*, Academic Press, Inc., New York, 1979.

[16] Maybeck, P. S., *Stochastic Models, Estimation, and Control: Volume 2*, Academic Press, Inc., New York, 1982.

[17] Maybeck, P.S., Hentz, K.P., "Investigation of Moving-Bank Multiple Model Adaptive Algorithms," *AIAA Journal of Guidance, Control, and Dynamics*, 1987, Vol. 10, pp. 90-96.

[18] Norden, Peter V., "Useful Tools for Project Management," in *Management of Production*, M. K. Starr, Editor., Penguin, Baltimore, Maryland, 1970, pp. 71-101.

[19] Parr, F. N. "An Alternative to the Rayleigh Curve Model for Software Development Effort," *IEEE Transactions on Software Engineering*, May 1980, pp. 291-296.

[20] Putnam, Lawrence H., "A General Empirical Solution to the Macro Software Sizing and Estimating Problem," *IEEE Transactions on Software Engineering*, July 1978, pp. 345-361.

[21] Watkins, Harry, *An Application of Rayleigh Curve Theory to Contract Cost Estimation and Control*, Master's Thesis, Naval Postgraduate School, Monterey, California, March, 1982.

Subscriptions, Monographs & Workshop Reports Order Form/Invoice

SHIP TO:	Return form to:
Name: ______	Military Operations Research Society
Organization: ______	101 S. Whiting Street, #202
Address: ______	Alexandria, VA 22304-3418
______	Phone: 703-751-7290
	FAX: 703-751-8171
Phone: ______	E-mail: morsoffice@aol.com

Please note: **All prices include freight for US orders only! For international prices please call the MORS office.**

PHALANX Bulletin subscription	1 year — $20	2 years — $35		
Military Operations Research (MOR) Journal subscription	1 year — $40	2 years — $75		
SAVE$$** Combined both subscriptions and **SAVE$$ *PHALANX* and *MOR* Journal	**1 year — $55**	**2 years — $100**		

DESCRIPTION		QTY	Freight	PRICE	TOTAL
Simulation Technology 1997, Sessions 1,2 & 3	(1988)			$ 10.00	
Human Behavior and Performance as Essential Ingredients in Realistic Modeling of Combat—MORIMOC II, Vol 1 & 2	(1989)			$ 55.00	
MORIMOC III	(1990)			$ 10.00	
MORS/ITEA *Emphasizing the "E" in T&E*	(1991)			$ 10.00	
COEA in the Acquisition Process & the Role of OR in Performing COEA	(1992)			$ 10.00	
Simulation Validation Workshop—SIMVAL II	(1992)			$ 10.00	
Emphasizing the Cost in COEA—MORS/SCEA Mini-symposium	(1993)			$ 10.00	
ITEA/MORS *How Much Testing is Enough?*	(1994)			$ 10.00	
Joint Requirements Oversight Council Process Workshop (JROC)	(1994)			$ 10.00	
How Much Testing is Enough?	(1994)			$ 10.00	
Simulation Data and Its Management Workshop (SIMDATAM)	(1993 & 1995)			$ 10.00	
MOR Analyst's Handbook, Vol I: *Terrain, Unit Movement, and Environment*			$ 5.00	$ 20.00	
MOR Analyst's Handbook, Vol II, Area 1: *Conventional Weapons Effects (Ground)*			$ 3.00	$ 5.00	
SAVE$$ Combine *Military OR Analyst's Handbook,* Vol I and Vol II, Area 1			$ 5.00	$ 25.00	
Warfare Modeling			$ 5.00	$ 35.00	
Sales TAX applicable on BOOKS delivered in Virginia only! There is no sales tax on freight or subscriptions.		VA Residents		4.5% TAX	
☐ CASH ☐ CHECK ☐ Purchase Order/Invoice ☐ OTHER		TOTAL ENCLOSED			$
CALL THE MORS OFFICE FOR CREDIT CARD ORDERS		TOTAL INVOICED			$

I hereby certify that amounts invoiced herein do not exceed the lower of (1) the contract price or (2) maximum levels established in accordance with Executive Order 11640, January 26, 1972. This invoice is due and payable upon receipt.

About our Authors

MODELING PRESENTATIONS AT NATIONAL CONVENTIONS

by Dean S. Hartley III

DEAN HARTLEY is Senior Scientist of the Center for Modeling, Simulation, and Gaming at the Department of Energy Oak Ridge facilities. He is a Director of both MORS and INFORMS and the Past Chairman of MAS. Dean's interests include analysis of historical combat data, verification and validation, and karate.

STATISTICAL VALIDATION OF A COMMUNICATIONS NETWORK SIMULATION

by Ann E. M. Brodeen and Malcolm S. Taylor

As senior project leader and investigator, Ann Brodeen is responsible for the application of statistical theory to the research, development, and evaluation of tactical computer science and information technology. She enjoys volunteering her time to teach quantitative methods at a neighboring elementary school.

Malcolm Taylor is a Fellow of the American Statistical Association, the Royal Statistical Society, and the Army Research Laboratory. He is Adjunct Professor of Mathematical Sciences and of Operations Research at the University of Delaware. He maintains an interest in fuzzy logic, quite apart from this article.

A DATA ANALYSIS OF SUCCESS IN OCS, THE USE OF ASVAB WAIVERS, AND RACE

by R.R. Read and L.R. Whitaker

Robert R. Read teaches probability, statistics and data analysis at the Naval Postgraduate School. His research is both theoretical and applied. It includes applications to numerous military operations and personnel studies including the generation of attrition rates for manpower studies. He is a member of MORS, IMS and ASA. He enjoys golf and skiing.

Lyn Whitaker teaches statistics in the Operations Research Department at the Naval Postgraduate School. Her research interests are in reliability, survival analysis including applications of maintenance and inspection policies, and in categorical data analysis. She is a member of MORS, IMS and ASA.

NON-MONOTONICITY, CHAOS AND COMBAT MODELS

by J.A. Dewar, J.J. Gillogly, M.L. Juncosa

Jim Dewar has worked at RAND for 17 years. Though his Ph.D. is in mathematics he has primarily worked in strategic planning. The idea for this work came to him on an airplane while he was reading James Gleick's book about Chaos.

Jim Gillogly, formerly at RAND, is a computer scientist at Mentat, specializing in system design and development, computer security, and C hacking. He is president of the American Cryptogram Association, coordinates a mailing list dealing with the medieval Voynich Ms., and sings drinking songs in period costume at Renaissance Faire while juggling.

Mario Juncosa, applied mathematician, seduced by the So. California climate and Herman Kahn's enchanted descriptions of RAND, left the Aberdeen Proving Ground's BRL during its relatively exciting period of the incipient computer age over 4 decades ago for RAND during its even more exciting epic era and research atmosphere with very interesting problems. Juncosa has also taught applied mathematics, physics, and statistics at various times and at various universities, consulted for the legal profession, and plays handball whenever he finds willing players.

FINAL-COST ESTIMATES FOR RESEARCH & DEVELOPMENT PROGRAMS CONDITIONED ON REALIZED COSTS

by Mark Gallagher and David Lee

Major Mark Gallagher is military-trained operations research analyst (USAFA B. S. in 1983, AFIT M. S. in 1986 and Ph. D. in 1992). A year ago, he transferred from the Office of the Secretary of Defense Program Analysis and Evaluation to US Strategic Command. In his aspirations to grow from a hobby farmer to a gentleman farmer, Mark is currently analyzing sheep production.

Educated at the University of Missouri (BSEE 1959), Brown University (Sc. M. 1961), and the Illinois Institute of Technology (Ph. D. 1963), David Lee was an Air Force officer, then as a civilian, Director of the Applied Mathematics Laboratory in the Air Force Aerospace Research Laboratories, Head of the Department of Mathematics and Computer Science at the Air Force Institute of Technology, and Director of the Operations Research/Procurement Planning Division in the Office of the Secretary of Defense. In retirement he is a busier consultant than his wife wishes him to be.

MILITARY OPERATIONS RESEARCH SOCIETY, INC.

COPYRIGHT ASSIGNMENT AGREEMENT

THIS AGREEMENT, made by and between ____________________, hereinafter ("Assignor(s)") and MILITARY OPERATIONS RESEARCH SOCIETY, INC., hereinafter Assignee).

Assignor(s) (*for U.S. Government employees: to the extent legally permissible*) hereby grants, transfers, assigns, and conveys to Assignee, its successors and assigns, the entire title, right, interest, ownership and all subsidiary rights in and to an article entitled ____

____

____(hereinafter the "Work"), including, but not limited to, the right to secure copyright registration of the Work and to any resulting registration in Assignee's name as claimant. Notwithstanding the foregoing, Assignor(s) reserves all proprietary rights other than copyright, such as patent rights; a royalty-free right to reproduce the Work; and the right to prepare derivative works based upon the copyright in the Work.

Assignor(s) hereby confirms that Assignor(s) owns the entire title, right and interest in the work, including the right to reproduce, prepare derivative works based upon the copyright in the Work, and distribute the Work, whether or not the Work constitutes a "work made for hire" as defined in 17 U.S.C. Section 201(b); Assignor(s) agrees that no rights in the Work are retained by Assignor(s) except as stated above. Assignor(s) agrees to execute any documents that might be necessary to perfect Assignee's ownership of copyrights in the Work and to registration.

Assignor(s) represents that the Work has not been copyrighted or published; that it is not being submitted for publication elsewhere; and, if the Work is officially sponsored by the U.S. Government, that it has been approved for open publication.

This Agreement constitutes the entire agreement between the parties hereto; this Agreement supersedes any prior oral or written agreement or understanding between the parties; and, **in the case of a Work Made for Hire, this Agreement has been signed by the Assignor(s)'s employer.** This Agreement shall only be effective if and upon the date that the Work is accepted by Assignee for publication in *Military Operations Research*, the Journal of the Military Operations Research Society or *PHALANX*, the Bulletin of the Military Operations Research Society.

ASSIGNOR(S)	**ASSIGNOR(S)**
By: ____________________	By: ____________________
Title: ____________________	Title: ____________________
Date: ____________________	Date: ____________________

MILITARY OPERATIONS RESEARCH SOCIETY, INC. (ASSIGNEE)

By: ____________________

Title: ____________________

Date: ____________________